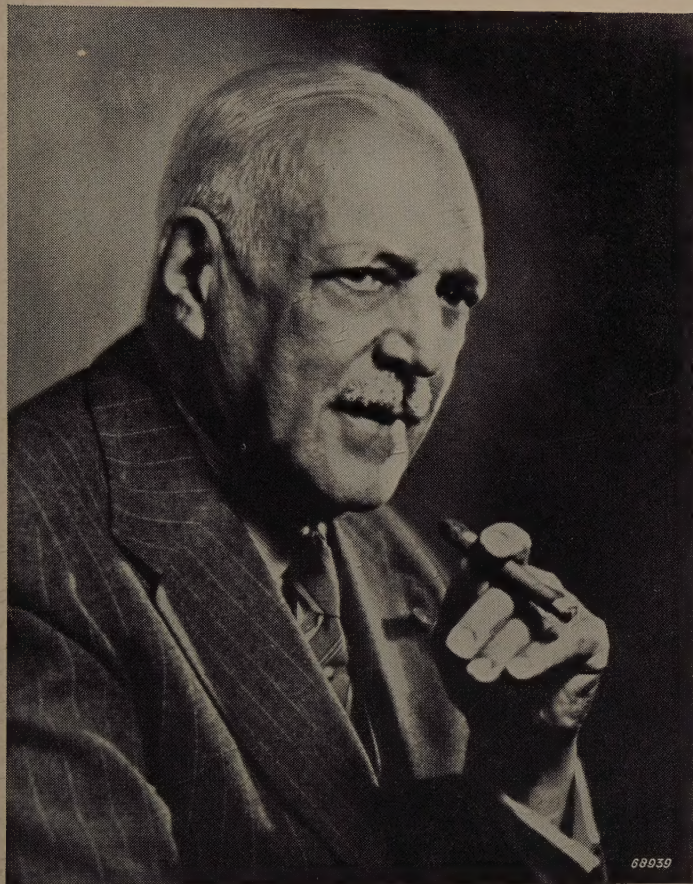


Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS

RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS



IN MEMORIAM Dr. h.c. A. F. PHILIPS

He was deeply convinced of the necessity of research, and the erection of a new building for the Physical Laboratory — founded in 1914 — was one of his first deeds after his brother resigned in 1922 and the management of the concern fell entirely into his hands. Although not a technician he on several occasions functioned as chairman of technical discussions in the laboratory when the development of important new technical products was to be promoted.

IN the early hours of Sunday, Oct. 7th, 1951, Anton Frederik Philips, born at Zaltbommel on the 14th of March, 1874, and one of the founders of the Philips concern, passed away at the age of seventy seven.

He was originally intended by his father, B. F. D. Philips, himself a banker, to occupy a position in the banking business. In 1895, at the age of 20, he was requested to assist his 16 years older brother, Gerard Leonard Frederik Philips, in developing the Philips lamp factory, founded in 1891, by undertaking the commercial part of the business, and not only did he succeed in bringing the young concern through its initial difficulties but he also very soon became, at his brother's side, the pushing power of this industry.

In 1939 Dr. Philips resigned as managing director, but in the years 1940-1945, in the U.S.A., he was still active in the interests of the concern and its overseas activities.

He did not assist personally at the festivities on the occasion of the 60-years' jubilee of the concern, May 1951, on the grounds of ill-health, but television made it possible for him to follow the events at home, which he did with lively interest. To the very last all questions concerned with the future of the Philips industries had his full attention.

The Philips personnel all over the world, and especially the staffs of the laboratories, deeply regret the passing of this remarkable man to whose indomitable initiative they are proud to owe their enviable sphere of activities.

THE EDITORS, PHILIPS TECHNICAL REVIEW

NEW ELECTRONIC TUBES EMPLOYED AS SWITCHES IN COMMUNICATION ENGINEERING

II. SWITCH TUBES

by J. L. H. JONKER and Z. van GELDER.

621.385.832:621.318.57:621.39

The authors continue their discussion of some experimental tubes designed for performing switching functions, in particular of tubes with which switching is brought about by the deflection of an electron beam. Among others, a tube has been constructed which can serve in this way as a multi-way switch.

Electromagnetic relays, such as are frequently used, inter alia, in line telephony, exist in a large variety. As far as the contacts are concerned, for instance, there are to be distinguished make contacts, break contacts and change-over contacts (see fig. 1 in a previous article¹), which can be operated in various numbers and combinations by a single excitation coil. Exclusively make and break contacts are one-way switches (mono- or multi-polar), but there are also multi-way switches, such as relays with change-over contacts, selectors and line finders, which connect one contact point to one of two or more other points. A distinction can also be made according to the switching speed: as a rule the switching speed has to be as high as possible, but in certain cases some delay is essential in order to ensure that the relay comes into action after another has performed its function.

In article I an electron tube was discussed which can act as relay with two contacts closed or opened simultaneously (bipolar symmetrical contact tube). Here it will be shown how tubes can be constructed for more complicated relaying functions.

In the tube just referred to, the control-grid voltage determines whether the electrons emitted by the hot cathode can reach a more positive electrode or not; this corresponds to the closing or opening of the contacts. Wider switching possibilities are afforded when the electrons are concentrated to a beam which, as in the case of a cathode-ray tube for an oscilloscope or for television reception, can

be directed by deflection and thus caused to act, as it were, as a switch arm. This type of tube has been given the name of switch tube. Such tubes have already been described in various publications and patents, and suggestions have been made for applying this principle for multifarious switching operations²).

A technical application is found in what is known as the distributor, a tube used in multiplex telephony systems for connecting a line (or radio link) to different channels in turn. In this tube the electron beam is deflected in such a way as to make contact in rapid succession with a number of electrodes arranged in a circle. For a multiplex connection two such tubes are needed, working synchronously.

The deflectable-beam type of switch tubes so far known have a number of serious drawbacks: owing to their large dimensions (they are much larger than a normal amplifying tube) they take up much space between the other elements and their size alone already makes them rather expensive; the high working voltage (some thousands of volts) is an undesired complication; the maximum permissible beam current (in the order of 100 μ A) is low and as a consequence the resistance of the external circuit is high, so that the time constant, which is usually required to be very small, is rather large.

When in a tube like the distributor the beam current is modulated with the aid of a control grid the output voltage is in anti-phase with the input voltage, and in many cases this is an objection.

¹) J. L. H. Jonker and Z. van Gelder, New electronic tubes employed as switches in communication engineering, I. Contact tubes, Philips Techn. Rev. 13, 49-54, 1951 (No. 3), here further referred to as I.

²) A similar tube was described in: Tj. Douma and P. Zijlstra, Recording the characteristics of transmitting valves, Philips Techn. Rev. 4, 56-60, 1939 (figs. 2, 3 and 4).

As explained in article I, however, by means of secondary emission — which is not employed in the tube referred to in note ²) — it can be so arranged that the voltage variations of the output electrode are in phase with those of the corresponding input electrode. Instead of bearing the character of a conducting switch arm, owing to the secondary emission, the beam then functions rather as a movable element capable of effecting contact between two adjacent electrodes.

The aim should therefore be to produce cathode-ray tubes — provided with secondary-emission contacts — in the size of normal amplifying tubes and suitable for the voltages and currents commonly used with them (e.g. some hundreds of volts and a few mA). A good step in the right direction was the introduction of the ribbon-shaped electron beam, which will now first be discussed.

Ribbon-shaped electron beam

In a contact tube such as described in article I there is no difficulty in producing a primary stream of electrons which, given a fairly large cross section, has an intensity of 1 to 2 mA and which can be passed through or blocked as desired with the aid of a control grid. The accelerating voltage required need be no more than say 300 V.

It is a different matter, however, when an electrostatically deflectable beam is required, which has to serve two or more contacts and, in order to avoid cross-talk, has to be focused on one contact at a time. A similar problem is encountered in cathode-ray tubes for oscilloscopes and for television receivers: for the sake of a sharply defined picture the spot on the fluorescent screen has to be small. Owing to the mutual repulsion of the electrons this is possible only when the current density in the beam is small (i.e. with low current strength) and the electrons have a high velocity (i.e. a high voltage). That is why oscilloscope and television tubes are made for currents in the order of 10 to 100 μ A and for voltages of 1000 to 25,000 V. Since the picture field has to be rather large these tubes are made with the familiar elongated shape.

An important factor arising in the problem of focusing is the space charge in the electron beam. The greater the space charge the more difficult it is to keep the beam focused, owing to the mutual repulsion of the electrons. In the formulae for calculating the space charge in a beam a number of quantities always appear in the following relation ³):

$$\frac{I}{A} \frac{l^2}{V^{3/2}}, \dots \dots \dots (1)$$

where I is the current strength, A the cross section of the beam, l the length of the beam between the last and the penultimate electrodes, and V the potential in the space between those two electrodes. The smaller the factor given by (1), the smaller is the space charge. This already points to one way of arriving at tubes having the desired higher currents and lower voltages, namely a reduction of l , which in our case is permissible since with switch tubes there is no need for the beam to cover any large field. Naturally this reduction in the length of the beam is also favourable for producing a smaller tube.

The reduction of l , however, is not sufficient to allow of the current I being raised to some milli-amperes and the voltage V being lowered to, say, 300 V without making the expression (1) and thus the space charge intolerably large. Some other improvement is therefore needed, which is found by departing from the usual circular cross section of the beam and giving it a flat rectangular shape, i.e. in the form of a ribbon. This is possible in view of the fact that in the case of a switch tube the beam has to be deflected in one direction only, so that high definition is required only in that direction. (This definition is necessary on account of the large number of contact systems which in a tube required to act, for instance, as a distributor have to be mounted side by side within a limited space.) This is contrary to the case of a tube used for an oscilloscope or television receiver, where high definition is required in two directions of deflection at right angles to each other, which in itself makes it necessary for the cross section of the beam to be circular (or square).

The ribbon-shaped electron beam has several advantages over a beam of circular cross section.

If we give the two beams the same definition in the direction of deflection, that is to say if the short side of the rectangle, d , is made equal to the diameter of the circular beam (*fig. 1*), then with the same current density the current in the ribbon-shaped beam will obviously be much greater than that in the circular beam.

Further, calculations prove that — again with equal current density — the adverse effect of the space charge (mutual repulsion of the electrons) is so much less in the case of a broad beam that the

³) K. R. Spangenberg, Vacuum tubes, MacGraw Hill Book Co., New York, 1948 (pp 443 et seq.).

strength of current can be increased two-fold to four-fold.

Thanks to these two advantages, with a ribbon-shaped beam currents of some milliamps can be obtained and the voltages need not be higher than 200 to 300 V.

A third important advantage of the ribbon-shaped beam is the simpler build-up of the electrode system. The apertures in the electrodes, forming together the "gun" which focuses the electrons, need only be accurately aligned in one direction (at right angles to the flat side of the beam). Thus it is pos-

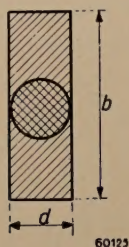


Fig. 1. Cross section of a circular electron beam (diameter d) and of a ribbon-shaped beam ($d \times b$). The latter beam is deflected to the left or to the right.

sible for the electrode system to be mounted between supporting plates of mica, as in normal amplifying valves, which greatly simplifies manufacture.

In another respect, too, advantage can be taken of the technique in construction of normal valves, by following the same principles for keeping the capacitances between the electrodes small. In the case of switch tubes this helps in obtaining a small time constant (short switching time).

Two further advantages of the ribbon-shaped beam are the following:

1. The focal length of a slit-shaped electronic lens, given the same potentials and field strengths, is only half that of a circular lens. Consequently the flat beam can be made shorter than the circular one, so that lower potentials are required.
2. In order to avoid cross-talk, the beam must be strictly focused on the one, desired contact system. This implies that the cathode must be sufficiently far removed from the lens to reduce the size of the "image" of the cathode to less than that of the contact system. Consequently the field strength at the cathode becomes rather small and possibly not sufficient current density can be obtained. Owing to the comparatively great width of the ribbon-shaped beam, which may be regarded as a large number of narrow beams in parallel, the density of current may be small without making the total beam current too small.

Tubes with one or more make and break contacts or change-over contacts

The tube illustrated in *figs 2 and 3* has two contact systems each comprising three electrodes ($h-r-a$ and $h'-r'-a'$, cf. *fig. 8* in article I). With the aid of the deflecting plates d_1-d_2 an electron beam

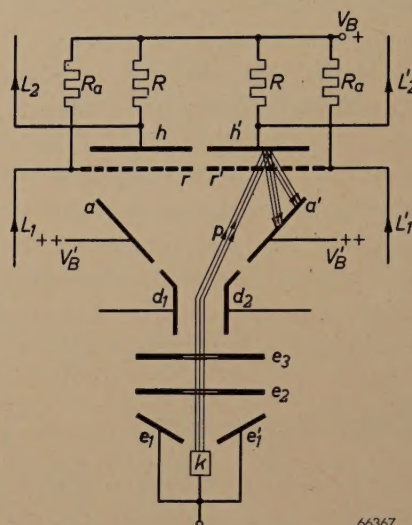


Fig. 2. Cross section of a switch tube with a make and a break contact. The cathode k supplies the electrons from which a ribbon-shaped beam p is formed by the electrodes e_1-e_1' , e_2 and e_3 . With the aid of the deflecting plates d_1-d_2 , this beam can be focused onto h or onto h' . $h-r-a$ and $h'-r'-a'$ are contact systems according to *fig. 8* of article I. A connection is made either between the lines L_1 and L_2 or between L_1' and L_2' .

can be directed upon either of the contact systems as required, thus closing either the contact $h-r$ or the contact $h'-r'$. In this way the tube can replace a relay having a make and a break contact, or, if r is connected to r' , a relay with a change-over contact. Owing to the grid-shaped input electrode

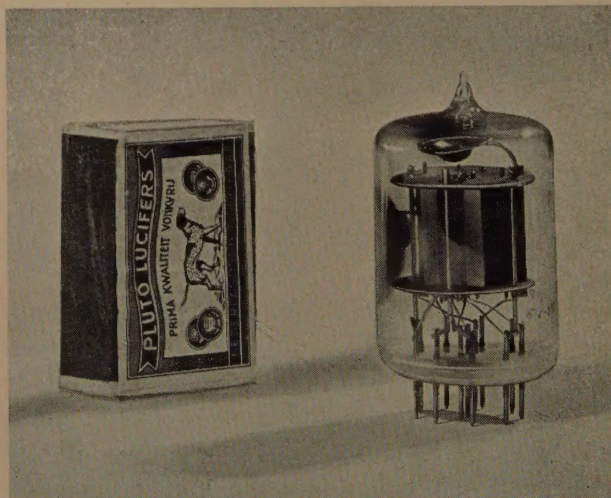


Fig. 3. Photograph of the tube represented in *fig. 2*.

used in the construction of the contact systems, as explained in article I, about 10 contacts can quite well be connected in series.

For the tube with only two contact systems it is not necessary that the beam should be particularly narrow in the direction of deflection, but even so it is advantageous to use a ribbon-shaped beam: the smaller the thickness d of the beam (in the direction of deflection), the more may the deflecting voltage (between d_1 and d_2) vary without causing the beam to diverge from the contact system upon which it is focused.

As shown in fig. 8 of article I, r was connected via a resistance to a point of a higher potential (V_B') than h . This is favourable for effecting a bilateral contact, such as explained for a similar case with reference to fig. 2e in article I. If a unilateral contact suffices then it is an advantage, as represented in fig. 2 (page 83), to feed r and r' with the same low voltage (say 70 V) as that applied to h and h' , because when the beam is directed upon h (or h') this does not in itself cause any change in potential at h (or h'). This means that in the case of telephone-current transmission the switching-over of the beam does not give rise to any click, and in the transmission of pulses there is no false signal.

If the width b of the ribbon-shaped beam is made large enough this beam can be made to cover a second, a third or more pairs of contact systems simultaneously, so that the tube can perform the function of a multipolar relay (with two or more make and break or change-over contacts).

By connecting the secondary-emission electrode h to the deflecting plates of one or more additional tubes it can be so arranged that the voltage variation brought about at h when this is struck by the beam is just sufficient to deflect the beam in the following tubes onto another contact system. This is equivalent to a relay contact closing or breaking the circuit of one or more other relays. This will be reverted to later.

Selector tube

The second type to be discussed is the selector tube, i.e. a switch tube which can connect one electrode to any other one of a large group of electrodes.

Fig. 4 is a sketch representing the electrode system. According to the amplitude of the voltage between d_1 and d_2 one of the output electrodes h_1, h_2, h_3, \dots is struck by the beam and owing to secondary emission that electrode assumes the constant potential of a screen grid r , which forms the one electrode to which each of the output electrodes can be "connected". The slots in the

screen grid r and the small thickness of the ribbon-shaped beam avoid the possibility of more than one output electrode h being in contact with r at the same time.

In front of the screen grid r is a suppressor grid g_3 , which is at zero potential; the wires of this grid are parallel to the plane of the drawing of fig. 4. This grid prevents any undesired transmission of secondary electrons to the deflecting plates. A screen s , likewise at zero potential, behind the electrodes h prevents electrons reaching the glass bulb.

As an application of the selector tube may be mentioned its use as a distributor in multiplex systems, reference to which has been made in the introduction. A sawtooth voltage between the de-

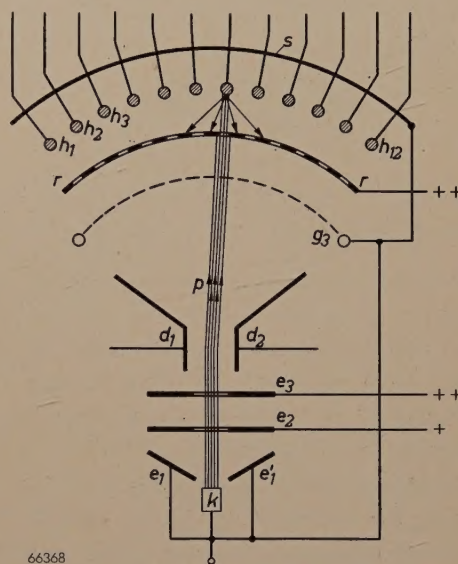


Fig. 4. Selector tube. The ribbon-shaped beam of primary electrons p is focused onto one of a large number of electrodes h_1, h_2, h_3, \dots which emit secondary electrons, whereby the electrode struck by the beam assumes the constant and high potential of the screen grid r . g_3 is a suppressor grid and s a screen, both at cathode potential, preventing electrons from reaching undesired places. The rest of the letters have the same meaning as in fig. 2.

flecting plates causes the beam to scan the contacts periodically. By means of a simple circuit, which cannot be gone into in detail here, a selector tube acting as distributor can be synchronized with another. The pulses of constant amplitude on the output electrodes h trigger in turn a repeater valve in each of the channels ⁴⁾.

A selector tube with six, or even with twelve contacts need be no larger than a normal radio valve (fig. 5).

⁴⁾ Cf. the channel-gate circuits described in: C. J. H. A. Staal, An installation for multiplex-pulse modulation, Philips Techn. Rev. 11, 133-144, 1949.

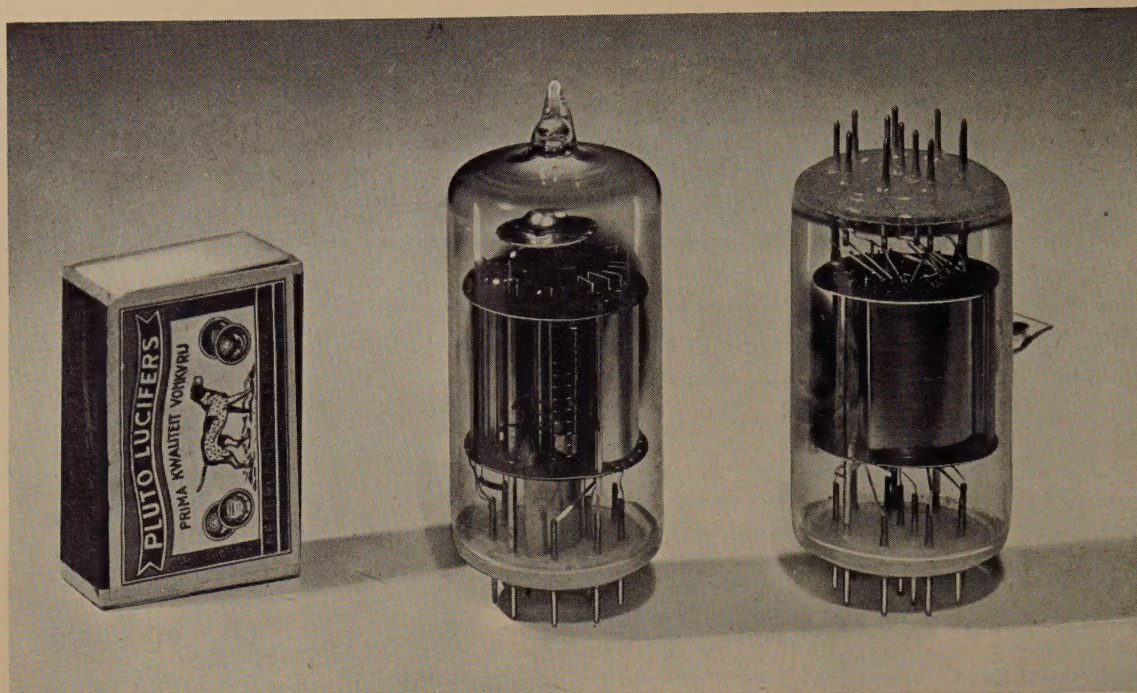


Fig. 5. Two selector tubes according to fig. 4. On the left one with six and on the right one with twelve contacts h .

Signalling tube

In telephony signalling is understood to comprise everything connected with the transmission of signals excepting the conversations themselves, i.e. signals for selecting or dialling, ringing, buzzing for "engaged", etc. These signals are in the form of pulses.

The contact tubes in question can, of course, also serve for transmitting signals with this special form of modulation, but nevertheless a switch tube has been designed with a system of electrodes as sketched in fig. 6, which is particularly suitable for this purpose. Fig. 7 is a photograph of two signalling tubes, one of which is fitted with a "Noval" base.

This tube, too, works with a ribbon-shaped beam, which can be deflected to the left or to the right with the aid of the deflecting plates d_1 and d_2 . The beam then enters a space enveloped by the grids r_1 (r_1') and r_2 (r_2'), the partition t and the plate n (n'); the grids and the partition have the same constant high potential V_B' (say 200 V); n and n' are the input electrodes receiving the signal pulses (with positive polarity). In fig. 6 the line 1 represents the beam deflected to the right in the case where the plate n' receives a signal with a positive voltage approximately equal to V_B' . After passing through the grid r_1' the primary electrons come into a space which is practically free of any electric field, so they shoot straight on and strike the electrode h' .

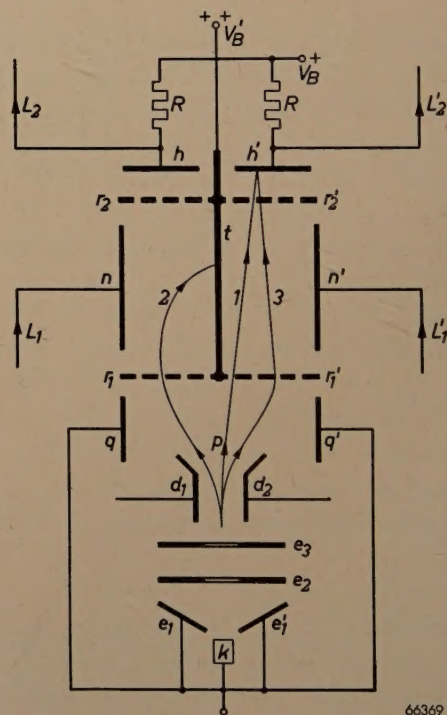
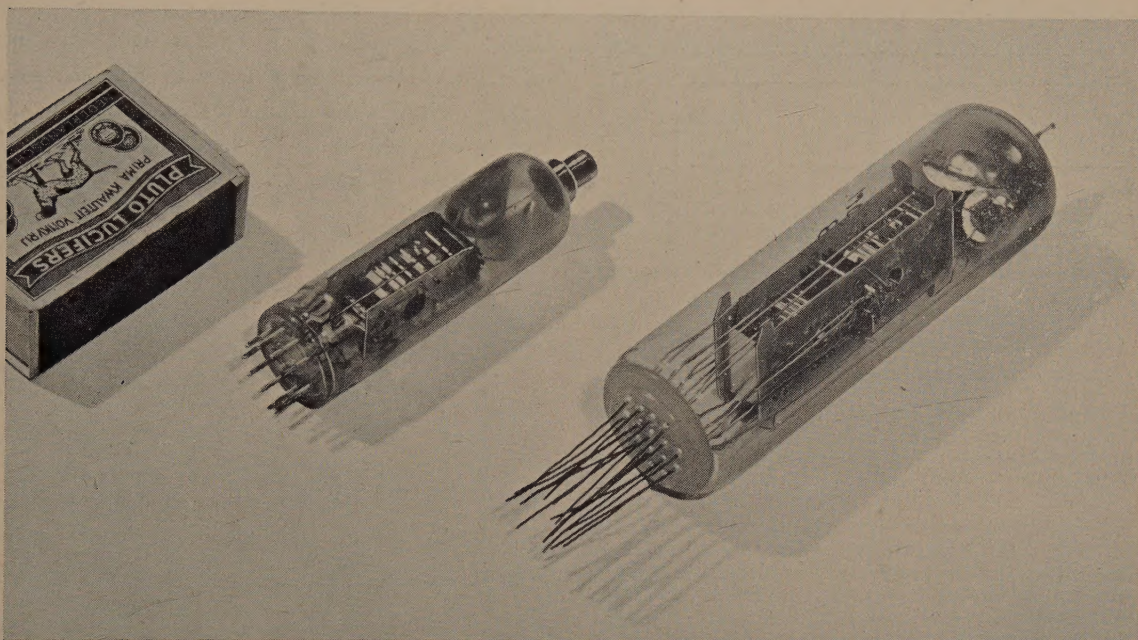


Fig. 6. Cross section of a signalling tube. For the meaning of k , e_1 - e_1' , e_2 , e_3 , d_1 - d_2 , p , h and h' , see fig. 2. The plates q - q' at cathode potential prevent the beam from being deflected too far. The collector t and the grids r_1 - r_1' and r_2 - r_2' are at the high potential V_B' . 1 is the path followed by the primary beam when this is deflected to the right and the plate n' (input electrode) has approximately the potential V_B' ; as the beam strikes the output contact h' the latter assumes the potential V_B' of the grid r_2' , and the same is the case when owing to a higher deflecting voltage the beam follows the path 3. The beam (deflected to the left) follows the path 2 if the incoming line L_1 and the input electrode n are at a low potential. The beam does not then reach h and the outgoing line L_2 does not undergo any change in potential.

The latter is thus caused to emit secondary electrons, which pass over to t and r_2' , and h' thereby assumes the high potential of r_2' , which forms the output signal.

In the absence of the input signal (or if the potential of the input electrode drops below a certain value) the situation is as indicated by 2 in fig. 6 (for the beam deflected to the left): owing to the low potential of n the beam is diverted away from n and strikes against the partition t , without reaching h .

fore, is analogous to the case of the selector tube, and essentially different from that of the switch tubes previously discussed where the voltage at the output electrode h follows that of the input electrode a . This implies that the signalling tube cannot transmit any arbitrary modulation of the input signal, but that on the other hand it is excellently suited for imparting pulses the amplitude of which is adjustable with the size of $V_{B'}$ and independent of the amplitude of the input pulses.



66133

Fig. 7. On the left a signalling tube according to fig. 6 with "Noval" base. On the right a double signalling tube incorporating two of the electrode systems illustrated in fig. 6 (but with one common cathode).

On either side is a plate, q and q' , at cathode potential. These plates prevent the beam from being deflected too far, so that in the event of the deflecting voltage being higher than necessary the primary electrons follow a path as represented by 3 in fig. 6. Thus, owing to the effect of the plates q and q' the deflecting voltage between the plates d_1 and d_2 is not at all critical and, moreover, the dimensions of the deflecting system can be kept small.

The plates n and n' have the character of control electrodes, rather than that of contacts. Here the "closing of the contact" means that the electrode h or h' , which previously had the potential V_B (say 70 V), assumes the potential $V_{B'}$ (say 200V), regardless of the voltage at n (or n'), except that it must have exceeded a certain value. This, there-

Contrary to the case of other switch tubes, there is practically no limit to the number of signalling tubes that can be connected in series.

This is to be explained as follows. A series connection of the contacts of three tubes according to fig. 2 is represented in fig. 8a, with an equivalent circuit in fig. 8b. When the three contacts are closed the resistances R_i of each contact are in series, and although R_i may be small with respect to R , a limit is therefore nevertheless set to the number of contacts that can be connected in series (maximum about 10). When, however, a number of signalling tubes are connected in series (fig. 8c) the equivalent circuit is as shown in fig. 8d: here a "closed contact" is represented by the connecting of an output electrode h via the contact resistance R_i ($\ll R$) to the voltage source $V_{B'}$, so that the resistances R_i are not connected in series and consequently any number of contacts of these tubes can be connected in series.

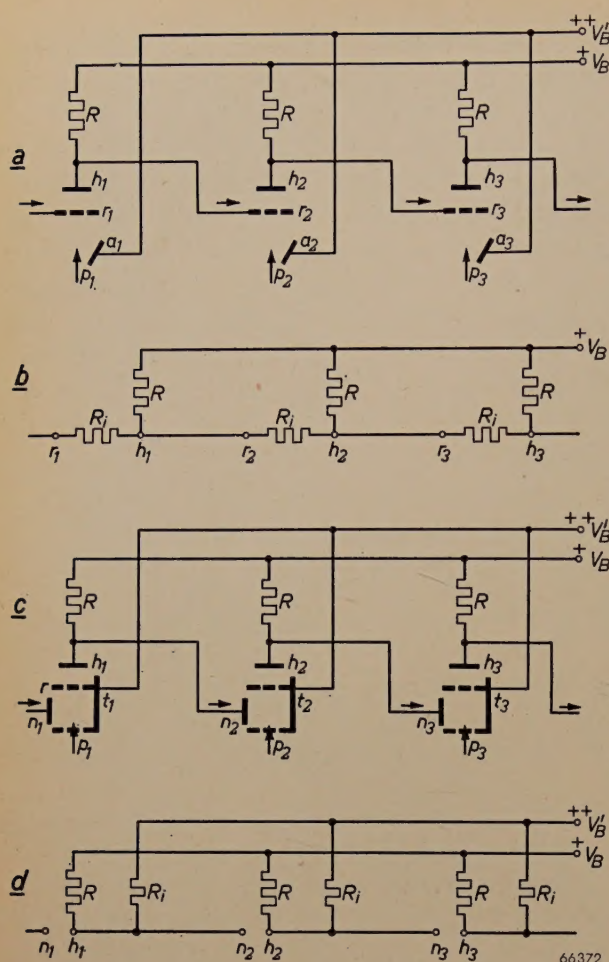


Fig. 8. a) Series connection of the contacts of three tubes according to fig. 2 (only the left-hand halves are drawn here). b) Equivalent circuit for the case where the three contacts are closed; the contact resistances R_i ($\ll R$) are thereby brought into series. c) Series connection with tubes according to fig. 6 (only the left-hand halves drawn). d) Equivalent circuit of (c) with the three contacts closed; here the contact resistances R_i ($\ll R$) are not in series.

Some switching-technical problems

Intentional delayed action

In the introduction it was already pointed out that sometimes it is necessary that certain (electro-magnetic) relays, simultaneously activated, should close their contacts in a predetermined sequence. For this purpose special constructions are employed with which a certain intentional delay is obtained, adjustable or fixed.

To attain the same object with switch tubes one may proceed as follows. In an auxiliary tube (fig. 9) a stream of electrons p_2 is generated and directed as a beam upon an anode a_2 . The deflecting plates of the auxiliary tube are connected to the corresponding plates of the tube whose action is to be delayed; this tube may be of the type shown in fig. 2 or fig. 6. The anode a_2 is connected via a resistor to a point of positive potential and

via a capacitor C to the deflecting plate d_1 . When the deflecting voltage of the other plate (d_2) is raised to a level whereby the beam is made to diverge from anode a_2 (the beam in the other tube then still strikes the electrode h) the potential of a_2 suddenly rises and as a consequence also the potential of d_1 is increased (both in the auxiliary tube and in the switch tube), thereby retarding the deflection of the beam in both tubes. This retardation is determined by the capacitance C and the resistances through which the charges can leak away.

If necessary a variety of electrode systems can be built into the auxiliary tube, with different RC times, so as to provide for different delay times. In principle it is also possible to mount the anode a_2 in the switch tube whose action is to be delayed and to arrange for a part of the (widened) beam to be directed upon it.

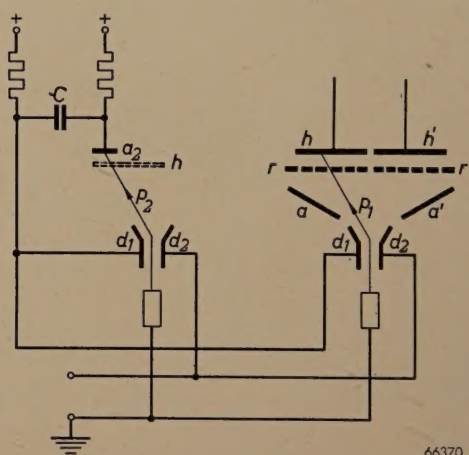


Fig. 9. On the right a switch tube according to fig. 2, the action of which is given a certain delay by means of the auxiliary tube on the left. When the voltage at the two right-hand deflecting plates d_2 is raised to such a level that the beams p_1 and p_2 are deflected to the right, instead of to the left, then upon p_2 leaving the anode a_2 the potentials of a_2 and (via the capacitor C) the two plates d_1 are suddenly increased, thereby retarding the deflection of the beams in both tubes. In the sketch of the auxiliary tube the output electrode h of the main tube has been drawn in broken lines so as to indicate that p_1 is still directed upon h when p_2 is just leaving the anode a_2 .

Replacement of several relays by one switch tube

Fig. 10a represents a certain combination of three monopolar relays which can be replaced by a single switch tube according to fig. 6. The deflecting plates d_1-d_2 , directing the beam to the left or to the right, take the place of the coil of relay 1, whilst the plates $n-n'$ in fig. 6 replace the coils of the relays 2 and 3 (the "contact" is only brought about when the potential of the latter electrodes is sufficiently positive).

The relay circuit can also be replaced by the tube represented in fig. 2, with r , r' , h and h' fed from one and the same voltage source. Relay 1 again corresponds to the switching-over of the beam. To raise the potential, for instance of h' , it is not sufficient to focus the beam on this electrode, but also the potential of r' has to be raised to the level desired for h' . The analogy of this in the relay circuit according to fig. 10a is that, in order to connect the lowermost point with the higher point on the right, not only relay 1 but also relay 3 has to be activated.

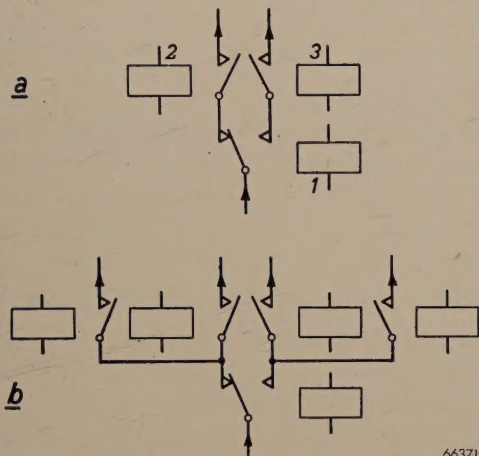


Fig. 10. The relay system *a* can be replaced by a tube according to fig. 6, and the system *b* by a similar tube with two contact systems side by side.

When there are two such groups of relays it can be arranged for the current passing through the coil of relay 1 of one group to be switched on and off by the contact of relay 2 or 3 of the other group. With two tubes according to fig. 6, for instance, a corresponding circuit can be built up in the manner already described, by using the output voltage of an electrode h of one tube as deflecting voltage between the plates d_1 - d_2 of the other tube.

If the beam is made wide enough to cover two contact systems placed side by side then one such a double tube can replace the five simple relays according to fig. 10b. If the whole of the electrode system were to be doubled (except for the cathode, which then has to produce two beams each serving two pairs of contact systems) then in this way, in a circuit similar to that of fig. 10b, even one tube could perform the functions of ten monopolar relays! It is obvious that this means a considerable saving in space.

It will be clear that the possibilities are not confined to the examples given here. Practice will have to show how many and what combinations can best be incorporated in one tube.

Summary. In this second article on "electronic switching" some experimental tubes are dealt with in which, as in the case of those already described, "contacts" are brought about by means of secondary emission (so that the output signal is in phase with the input signal), but with the beam of primary electrons guided in the desired direction by electrostatic deflection, thus giving these tubes the character of a change-over switch. The use of a ribbon-shaped beam, having a cross section large in one direction but small in the other, has a number of advantages: the strength of the beam current can be raised to several milliamps, the voltages required need not be higher than 200 to 300 volts, and the dimensions of the tubes can be kept small (about the size of normal radio receiving valves).

There are dealt with in turn a tube with one or more make and break contacts (or change-over contacts), a selector tube with 6 and one with 12 contacts, and a signalling tube (which can transmit only pulses; the amplitude of the pulses transmitted is independent of that of the input pulses, provided that the latter exceed a certain value). If desired an adjustable delay can be introduced in these tubes. Some examples are given to explain how one single switch tube can replace a number of monopolar relays.

NEW VIEWS ON OXIDIC SEMI-CONDUCTORS AND ZINC-SULPHIDE PHOSPHORS

by E. J. W. VERWEY and F. A. KRÖGER.

537.311.3:535.371

Just as a few decades ago great advances in the electrotechnical industry followed the introduction of electron tubes such as the diode and triode, the cathode-ray tube and the thyatron in electric engineering, so now great possibilities can be envisaged by the application of specific properties of solids. All kinds of new materials with special properties or combinations of properties are being introduced: Semi-conductors play a large part in blocking-layer rectifiers, as basic material for resistors, photo-conductive layers, etc.; fluorescent materials are being used on a large scale in illumination engineering, television, radar, etc.

In view of the great practical importance of these materials there is every reason to study them intensively also from the theoretical aspect and to attempt to gain a deeper insight into their behaviour.

It has long been known that important physical properties of solids are often governed by certain "impurities". As examples may be mentioned the conductivity of semi-conductors and the optical properties of fluorescent substances. In many cases the mechanism whereby the "impurities" exercise their influence is still obscure, while in other cases although some idea can be formed of this mechanism it is extremely difficult to prepare a substance with the desired content of impurities.

With regard to some points, however, a deeper insight has recently been gained, and new possibilities for a better mastery of the preparation of materials with desired properties have thereby been gained.

In this article semi-conductors will be dealt with first. Following upon a discussion of some ideas of older date, it will briefly be explained how new ideas have led to the "method of controlled valency". Then some new views will be discussed which have been gained by consideration of the zinc-sulphide phosphors, from which it will appear that there is a great analogy in the lines of thought followed in both these fields.

Semi-conductors

Semi-conducting elements

The first materials whose behaviour under the influence of impurities came to be fairly well understood were the semi-conducting elements silicon and germanium, thanks to extensive investigations carried out by, among others, Lark-Horowitz and Johnson¹). It appears that when

an element in the third group of the periodic system, for instance boron, aluminium, gallium or indium, is admixed with Si or Ge a semi-conductor is usually obtained which behaves as if its conductance were due to positive charges (a "hole" semi-conductor). If the impurity is an element in the fifth group of the periodic system, such as nitrogen, phosphorus, antimony or arsenic, usually a semi-conductor is obtained whose conductance is due to negative charges.

This behaviour is to be explained by the obvious supposition that the foreign atoms occupy a normal place in the lattice. Considering, for instance, the case of germanium containing antimony, it can be imagined that four of the five valency electrons of the Sb atom take part in the chemical bonding with the four closest neighbours. (Ge and Si both have four valency electrons and crystallize into the so-called diamond structure, where each atom has four immediate neighbours.) The fifth electron is, as it were, superfluous and can be regarded as an electron which in a medium (germanium) with dielectric constant ϵ moves in the Coulomb field of a positive Sb^+ ion.

Owing to the high value of ϵ it requires only an energy of some hundredths of an electron-volt to release such an electron from the Sb^+ ion, after which it is comparatively free to move about in the crystal. Thus, even in very small quantities such an impurity causes a considerable increase in the conductivity of the crystal.

The above-mentioned hole-conductance occurring in the case of contamination with an element of the third group of the periodic system can be explained in a similar manner. The replacement of a Ge or Si atom by, for instance, Al, gives rise to a deficit in electrons. It can be assumed that

¹) See, for instance, K. Lark-Horowitz, *El. Eng.* **68**, 1047, 1949; W. Shockley, *Electrons and holes in semi-conductors with applications to transistor electronics*, Van Nostrand, New York, 1950.

an electron from a Ge atom in the immediate vicinity of the Al atom has the tendency to form, together with the Al atom, a negative ion. Of course it is a matter of indifference which of the immediate Ge neighbours releases an electron. In the position from which at a certain moment an electron is missing there is said to be an electron "hole".

This "hole" behaves as a positively charged particle which, in a medium with dielectric constant ϵ , moves in the Coulomb field of the negative Al⁻ ion. As in the first case considered, the binding energy of this "positive charge carrier" is very small. After being released from the Al⁻ ion it can move about in the crystal practically unhindered, so that again there is a contribution towards the conductivity.

Apart from this superficially sketched theory of the origin of conductivity in Si and Ge there are also detailed theories by which it is possible to explain the dependency of conductivity upon temperature and upon the concentration of the impurity as found experimentally. By measuring the conductivity of materials to which known quantities of foreign atoms have been added, it is possible to calculate, with the aid of these theories, the specific resistance of ideal pure Si or Ge. It is then found that at room temperature this specific resistance should be respectively 10^6 and 10^2 Ωcm . When these values are compared with those measured with samples of Ge or Si which have been purified by the most rigorous methods, it appears that the specific resistance of the materials prepared in this way is still, respectively, a factor 10^4 and a factor 2 too small. From this it can therefore be concluded that these materials still contain a certain small amount of impurities which cannot be removed; this concentration is so slight that it is usually impossible even to determine their nature.

Non-stoichiometric semi-conductors

In addition to the semi-conducting elements mentioned there is a large group of semi-conducting compounds, many of which belong to the metal oxides or metal sulphides. Of the oxidic semi-conductors the non-stoichiometric semi-conductors have been known longest²⁾. Although the electrical properties of these semi-conductors are not governed by impurities in the sense of "foreign atoms", a few words will be devoted to them here. The fact is that the mechanism of conductance in these

substances is analogous to that in semi-conductors prepared on the basis of the method of controlled valency to be described below.

The non-stoichiometric semi-conductors owe their conductivity to a departure from the simple, stoichiometric, relation of the component elements. This deviation may consist in the occurrence of an excess of metal or an excess of oxygen. Generally speaking, an excess of ions of one kind is accompanied by lattice flaws; further, non-stoichiometric compounds are generally found in compounds of metals with varying valency. For example, in the case of Cu_2O with an excess of oxygen there must be a number of Cu^+ places unoccupied, whilst, for the electrical neutrality of the lattice, an equivalent number of Cu^+ ions must have changed into the bivalent state. Schematically the situation may be represented by $(\text{Cu}_{2-2\delta}^+ \text{Cu}_\delta^{2+} \square_\delta) \text{O}^{2-}$, where δ is the excess of oxygen in atomic percentage and \square represents a vacant site in the lattice. Given sufficiently low temperatures it can be expected that these Cu^{2+} ions will be found next to the vacant Cu^+ sites, since, for Cu^+ ions in the immediate neighbourhood of the vacant Cu^+ sites, less energy is required to remove an electron from a Cu^+ ion, on account of the positive charge being lacking (see *fig. 1a*). At higher temperatures an electron can also be released from one of the many ions farther removed from the vacant site. This electron may then again convert the original Cu^{2+} ion into a Cu^+ ion. It is then said that the electron "hole" is removed from the immediate vicinity of the empty place in the lattice through thermal dissociation. Provided it is removed far enough the electron "hole" is able to move about in the lattice practically unhindered, since the various positions at a distance from the empty place in the lattice correspond approximately to one and the same energy³⁾ (see *fig. 1b*). It is in this manner that the non-stoichiometric Cu_2O receives its conductivity, and it is also seen that the conductivity increases greatly with rising temperature.

The practical utility of non-stoichiometric semi-conductors is strictly limited. In the first place it is almost impossible to produce a homogeneous, stable, non-stoichiometric semi-conductor with a particular conductivity value. In the second place the deviations from the stoichiometric state which can be reached are as a rule small. This implies also that the conductivity is limited in its variability. These limitations arise from the lack of stability in a lattice with many lattice flaws.

²⁾ See, for instance, E. J. W. Verwey, Philips Techn. Rev. 9, 47, 1949.

³⁾ J. H. de Boer and E. J. W. Verwey, Proc. Phys. Soc. 49, extra part, 59, 1937.

It will be shown how it has been found possible to produce oxidic semi-conductors in a manner whereby these difficulties are avoided and the conductivity is brought about, as in the case discussed above, by ions of the same element but with different valency occupying crystallographically equivalent places in the lattice.

The better conducting nickel oxide obtained in this way is of the type of non-stoichiometric semi-conductors discussed above and has the same drawbacks: it is not of practical utility owing to lack of stability. But an oxide containing Ni^{3+} in addition to Ni^{2+} ions can now be obtained in quite a different way, by admixing a certain quantity



Fig. 1. Two-dimensional model of an M_2O lattice with an excess of oxygen. The regular pattern characterizes a lattice in which the number of positive ions (denoted by +) is twice the number of negative ions (which obviously carry a double charge and are denoted by O). The vacant lattice site (surrounded by the broken line) expresses that a positive ion is missing. To ensure the electric neutrality of the lattice the number of double-charged positive ions (++) must be equal to the number of vacant sites (hence Cu^{2+} ions in the case of Cu_2O). Given a sufficiently low temperature, the electron "hole" corresponding to such an ion lies in the immediate vicinity of the vacant site in the lattice (a). Once the electron "hole" is removed from the vacant site in the lattice at higher temperatures, it is practically free to move about in the whole of the lattice (b).

The method of controlled valency

The new method of producing semi-conductors, indicated by Verwey, Haaijman and Romeyn⁴⁾, will be described by taking the case of nickel oxide as an example.

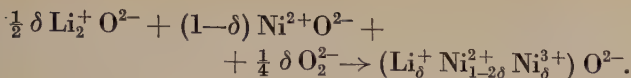
When NiO is prepared by heating NiCO_3 in an oxidizing atmosphere (say air) at about 1200°C , the oxide is obtained in the form of a green crystalline mass with a high specific resistance (up to $10^{12}\ \Omega\text{cm}$). This resistance is too high for practical purposes; better conductivity can be expected if in addition to the bivalent Ni ions present in the NiO compound there are also Ni^{3+} ions in the crystal lattice, and this can be arranged: when NiCO_3 is dissociated in air at a low temperature (about 500°C) a nickel oxide is obtained which has a small excess of oxygen and thus contains some trivalent nickel ions. The presence of these Ni^{3+} ions is disclosed by the black colour of the oxide.

of Li_2O with the NiO (in practice the two carbonates are mixed) and again heating this mixture to 1200°C in air. The product is a black oxide with low specific resistance. This oxide is a homogeneous crystalline mass, the composition of which is found by chemical analysis to be



the amount of Ni^{3+} being equivalent to the amount of Li added as impurity, at least as long as the latter amount is kept within certain limits. X-ray diffraction shows that the crystal structure of this black oxide is the same as that of NiO ; only the dimensions of the elementary cell have become somewhat smaller. From the above formula it follows that in the preparation of this oxide an amount of oxygen must have been absorbed from the air, whilst a certain amount of Li_2O is taken up in the lattice of the NiO , so that the final product is again an oxide with a minimum of vacant sites in the lattice. The reaction taking place may be written as follows:

⁴⁾ E. J. W. Verwey, P. W. Haaijman, F. C. Romeyn and G. W. van Oosterhout, Philips Res. Rep. 5, 173, 1950 (No. 3).



For $\delta = 0.1$ the specific resistance is about $1 \Omega\text{cm}$, a factor 10^{12} lower than the specific resistance of pure NiO (see *fig. 2*).

The artifice applied here in using lithium has a very general significance, as may be gathered from the following.

If the changing of the valency of some Ni ions were to be brought about by an excess of oxygen, then at the same time there would be some unoccupied Ni sites in the lattice. If we were to aim only at introducing Li_2O into the NiO lattice this would necessarily be accompanied by unoccupied O sites (or interstitial Li ions). Actually in our case neither of these two processes takes place separately. The process which does take place may be regarded as a combination of the two, giving the above-mentioned reaction and without leaving any vacant sites in the lattice.

This leads to the conclusion that the apparent preference for a situation without vacant sites in the lattice can be utilized for bringing about desired changes in valency by the admixture of suitable impurities. This is what the authors⁴) call the method of controlled valency.

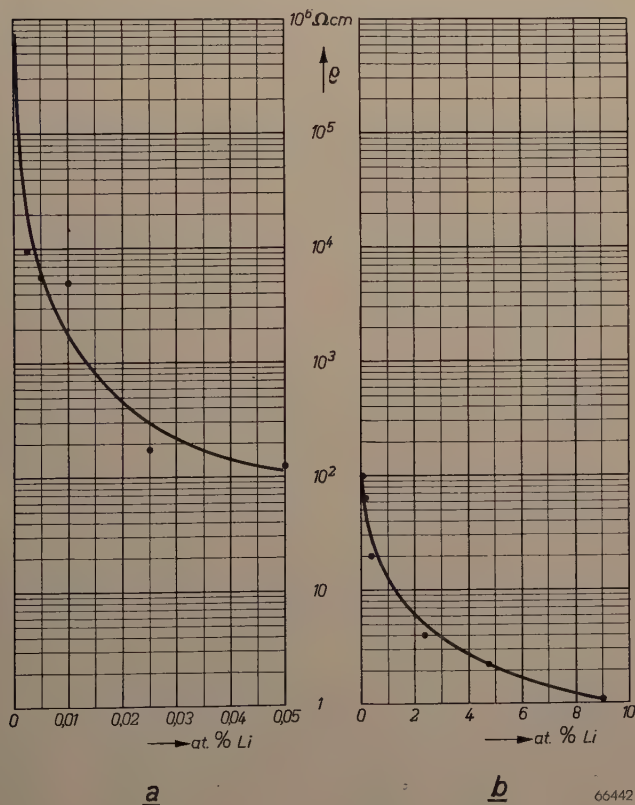
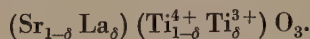


Fig. 2. Specific resistance of $(\text{Li}_\delta\text{Ni}_{1-\delta})\text{O}$ as a function of the Li content. The atomic percentage of Li varies between a) 0% and 0.05%, b) 0.05% and 9%.

For this method to be successful, certain conditions have to be fulfilled, which in our case of NiO mixed with Li may be formulated as follows. In the first place the Li must have a more stable valency than Ni. In the second place the ion radius of Li^+ must not differ much from the radius of Ni^{2+} , so as to permit the incorporation of Li^+ ions in the NiO lattice. These two conditions are indeed satisfied. The energy required to release an electron from an Li^+ ion, so as to change that ion into an Li^{2+} ion, amounts to 75.3 electron-volts, which is much greater than the energy required to turn a bivalent Ni ion into a trivalent one, namely 35 eV. The radii of the two ions Li^+ and Ni^{2+} are both 0.78 Å.

The expression given above for the reaction is based upon the assumption that a deficit of electrons in NiO leads to the formation of trivalent Ni ions and not to the formation of monovalent O ions. A closer theoretical study proves this assumption to be justified.

Practice has already shown that the method of controlled valency can be applied to a large number of compounds if only the suitable impurity is chosen. This makes it possible for the conductivity of these compounds to be adjusted within certain, wide, limits. Usually the maximum soluble quantities of the impurity are much smaller than in the case of NiO, where as many as 30% of the sites of metal ions can be occupied by Li^+ ions. In all these compounds a number of cations of the original lattice are replaced by foreign ions of a different charge, which, given an equivalent number of cations in the basic material, bring about a change in valency, so that the electrical neutrality of the whole crystal is preserved. The foreign ion may also occupy a site in the lattice which is not crystallographically identical to the sites occupied by the ions with variable valency. This is illustrated by the example:



In discussing the phosphors an example will be taken where the valency of the cations is changed by replacing some of the anions by foreign anions with a different charge.

Of the semi-conductors produced by the method of controlled valency there is one which has already found practical application as a resistor with negative temperature coefficient.

Zinc-sulphide phosphors

We shall now pass on from semi-conductors to zinc-sulphide phosphors, but first we would

briefly recall what has hitherto been known about the nature of the centres of fluorescence where the fluorescence is assumed to originate.

Purely stoichiometric ZnS does not show any visible fluorescence when irradiated (excitation) with ultra-violet rays or bombarded by electrons. The excitation energy is then either imparted to the crystal lattice in the form of heat through some process of dissipation or else radiation takes place in the ultra-violet. The incorporation of impurities in the ZnS lattice enables the ZnS crystal, after excitation, to radiate a large part of the excitation energy in the form of light. Hence the name "activators" sometimes given to these impurities⁵⁾, which may consist, for instance, of Ag, Cu, Au, or Zn. By mentioning Zn as an impurity in a ZnS lattice we have taken the conception of an impurity very broadly. When speaking of ZnS as having Zn as an "impurity" or being activated by Zn we mean that the ZnS has an excess of Zn.

As to the form in which the activators are absorbed, the general belief was that they are taken up in the lattice as atoms, occupying interstitial sites. According to other investigators they turn into an ion by releasing an electron, which in the ionic form occupies a normal cation site. In the latter case it was assumed that an equivalent number of vacant anion sites are formed which can just be occupied by the electrons released.

It will now be explained how more recent investigations have led to these ideas being revised.

ZnS phosphors are made by heating precipitated ZnS in a suitable atmosphere, for instance H_2S . In order to promote crystallization a so-called flux is added, which is subsequently removed by washing. Then, according to the colour of the fluorescence required, very small quantities of the elements Ag, Cu, Au are added as activators. In the case of a product having an excess of Zn one sometimes speaks of self-activated ZnS. It has now been found that traces of Cl, Br or I play an essential part in the formation of centres of fluorescence: a fluorescent material can only be produced if the flux contains one of these halogens or the atmosphere contains HCl or HBr⁶⁾.

For instance, pure ZnS heated in an atmosphere of H_2S and HCl shows a blue fluorescence, the intensity of which increases with the HCl content of the atmosphere (see fig. 3). Thus it appears

that Cl is very closely related to the formation of the centres. It is fair to assume that a number of S^{2-} ions are replaced by Cl ions, so that at the same time an equivalent number of Zn^{2+} ions change into the monovalent state. It can indeed be proved that the number of centres introduced into the non-activated ZnS is about equivalent to the amount of Cl absorbed in the lattice.

The concentration of the centres can be determined from the phenomenon of the "saturation" of the fluorescence. This phenomenon implies that when employing an exciting beam of electrons of great current density (an electron beam is used for the excitation because it is difficult to produce a sufficiently intensive ultra-violet radiation) the intensity of the fluorescence no longer increases linearly with the electron current but approaches a constant value. This saturation intensity is proportional to the concentration of the centres.

Similarly, also in the case of the phosphors activated with Ag, Cu or Au it can be assumed that the activator is taken up in the form of its halide and not, as was formerly believed, in the atomic form. Since the colour of the fluorescence is determined exclusively by the activator used,

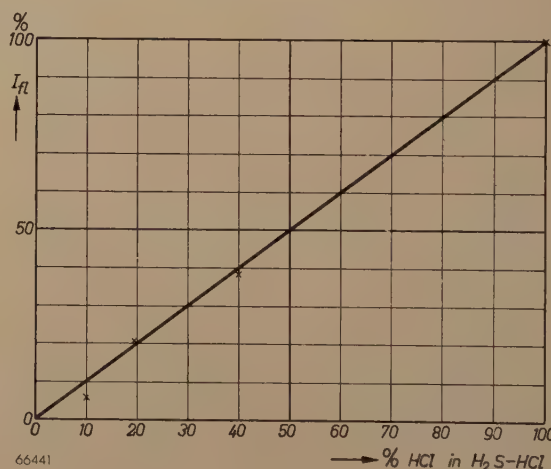


Fig. 3. Intensity of the blue fluorescence of ZnS.ZnCl as a function of the composition of the atmosphere in which the product is prepared. The curves relate to fluorescence after excitation by rays of a wavelength of 2537 Å at -180°C .

and not at all by the halogen used in the preparation of the phosphor, it can be assumed that the halogen ions are not present in the centres of fluorescence but that, in analogy with the semi-conductors, the halogen ions serve exclusively to keep the crystal electrically neutral (compensation of charge) and to prevent the formation of vacant sites in the lattice. This may also be expressed by saying that in this case we have to do with the ordinary formation of a mixed crystal ZnS-AgCl. The correctness of this assumption is proved by the following.

⁵⁾ See, for instance, F. A. Kröger and W. de Groot, Philips Techn. Rev. 12, 6, 1950 (No. 1).

⁶⁾ F. A. Kröger and J. E. Hellingman, J. Electrochem. Soc. 93, 157, 1948. See also footnote ⁷⁾.

It is also possible to prepare fluorescent zinc-sulphides in a manner whereby the charge-compensating function is taken over by a trivalent cation. In the neutralization of the charge introduced by the activator ($\text{Zn}^{2+} \rightarrow$ for instance Ag^+) the replacement of an S^{2-} ion by a Cl^- ion serves in the same way as the replacement of a Zn^{2+} ion by a trivalent cation⁷). For instance, ZnS containing as impurity exclusively Al or Ga shows blue fluorescence in a frequency band absolutely identical to the band in which the frequencies of the fluorescent light from self-activated ZnS lie. Also products containing Ag, Cu or Au together with Al, Y, Yb, La or Gd show fluorescence in the blue, green or yellowish green, which again is absolutely the same as the fluorescence from the corresponding phosphors prepared with halogen ions instead of trivalent cations⁸).

Thus it is seen that the fluorescence depends solely upon the activator used and not upon whether halogen ions or trivalent cations are used. Neither has the choice of any special halogen ions or trivalent cations any influence.

This applies with a certain restriction. Some trivalent cations, such as Ga, In, Sc, Pr or Ce, are in themselves capable of performing the function of activators. As a consequence, in the fluorescence from phosphors prepared with these elements there appear new fluorescence bands in addition to those ascribed to the actual activator. This complication will not be dealt with here⁸).

From the foregoing it can be concluded that the halogen ions or trivalent cations take no part in the actual process of fluorescence and, therefore,

cannot be considered as belonging to the centres. Apparently the centres of fluorescence consist of a monovalent activator ion with its surroundings of sulphur ions.

On the basis of extensive experiments, the description and discussion of which would lead us too far here, it can be plausibly concluded that the action of the activator ion lies mainly in its different (smaller) charge reducing the ionization energy of a neighbouring S^{2-} ion. To gain a more complete insight into the process, however, account has to be taken not only of the electrostatic forces but also of the polarization effects, repulsive forces and van der Waals' attractive forces. These are in fact responsible for the differences in fluorescence found when ZnS is activated with different elements.

These new conceptions not only form an important contribution towards the theory of fluorescence but at the same time present the possibility of producing entirely new phosphors with the aid of trivalent cations, such as would not be possible by the incorporation of Cl^- .

Summary. Following upon a brief discussion of the semi-conducting elements silicon and germanium and of non-stoichiometric oxidic semi-conductors, the "method of controlled valency" is described by taking nickel oxide as an example. This method consists in bringing about desired changes in valency by the admixing of suitable "impurities". There is a great similarity in the mechanism of conductance between non-stoichiometric semi-conductors and other semi-conductors prepared by the method of controlled valency. Compared with the former, the latter group however possess great advantages, which are to be ascribed to the lack of thermal stability of non-stoichiometric compounds. This lack of stability is due to the fact that in non-stoichiometric compounds there are always vacant sites in the lattice.

Somewhat similar considerations apply to zinc-sulphide phosphors. After the part played by halogen ions in the formation of these substances had been realized, a better picture could be formed of the fluorescent centres: the centres are formed by an activator ion surrounded by sulphur ions. The accuracy of the hypothesis developed finds confirmation in the possibility of having the part played by the halogen ions taken over by trivalent cations. This opens the way to the production of entirely new phosphors.

⁷) The possibility cannot be excluded that very small quantities of activator are taken up in the lattice without the presence of Cl, Br or I or of trivalent cations. The concentrations of the activator which can be introduced into the crystal when one of these ions is used in the preparation of the product are, however, so much greater that this fact may safely be ignored in our considerations.

⁸) F. A. Kröger and J. Dikhof, *Physica* **16**, 297, 1950.

COLLIMATING X-RAYS IN BEAMS OF VERY SMALL DIVERGENCE AND HIGH INTENSITY

by J. A. LELY and T. W. van RIJSSEL. 537.531:535.312:539.26:621.386

When X-rays fall upon a polished surface at a very small angle total reflection may take place. Use has been made of this principle in constructing an "X-ray lens" and even an X-ray microscope. It also enables one to make — as an aid in the X-ray diffractive examination of proteins and suchlike substances — a simple collimating system producing an X-ray beam with very small divergence, a relatively high intensity and very little scattered radiation.

X-ray diffraction

The examination of crystalline substances with the aid of X-ray diffraction is indispensable for the identification or chemical analysis of such substances and for the investigation of crystal structure, texture and many other properties. Several articles on this subject have already appeared in this Review¹⁾, so that there is no need to go into the theory of X-ray diffraction here. We would recall, however, the well-known Bragg's equation, which reads:

$$2d \sin \theta = n\lambda, \quad (1)$$

where d is the distance between a system of parallel lattice planes in a crystal, θ the angle between the incident ray and those planes, λ the wavelength of the rays and n a whole number giving the order of the reflection. Reflection takes place only when the angle θ satisfies the condition given in the formula; a ray striking the lattice planes at any other angle is not reflected. It is to be noted that in this connection it is always the angle between the incident ray and the reflecting lattice plane which is used, and not the angle between the direction of the ray and the normal to the reflecting plane, as is usually taken in optics. The scattering angle of the X-rays in the case of reflection is therefore equal to 2θ .

When employing the $\text{CuK}\alpha$ rays emitted with a wavelength of 1.54 \AA by a copper anticathode (copper is the material frequently used for X-ray diffraction work; molybdenum and tungsten are also used, but then the rays are of a shorter wavelength) and with spacings in the order of a few \AA such as occur in normal crystals, one finds for the scattering angles of the reflected rays values lying

between approximately 10° and 180° . Particularly the angles between 10° and 80° are of importance.

Divergence of the X-ray beam

Disregarding special focusing systems, in order to obtain diffraction angles which are defined with sufficient accuracy it is as a rule essential to have a narrow X-ray beam. This is obtained, for instance, by confining the rays emerging from the focus of the X-ray tube within certain limits fixed by means of two diaphragms set up some distance apart. The narrower the diaphragms, the less is the divergence of the X-ray beam, but at the same time also its intensity is reduced. When applying the Debye-Scherrer method, whereby the diffraction of the X-ray by a finely distributed crystalline powder is investigated (all kinds of lattice planes in the separate crystals then occur partly in the reflecting state), one finds that the reflections are rather weak. If in that case an accurate definition of the diffraction angle is desired, then the exposure has to be rather long (some hours) to be able to record all the reflections. Recording is mostly done on a photographic plate or film, but nowadays a Geiger-Müller counter is increasingly used.

Scattered radiation

In addition to the X-rays limited by the diaphragms and reflected by the lattice planes in the crystal, some scattered rays also fall on the photographic plate or film.

This scattering is caused by the edges of the diaphragms and by the molecules of the air. It can be suppressed for the greater part by taking suitable measures (rounding off the edges of the diaphragms, enclosing the primary X-ray beam in front of and behind the specimen in a narrow cylinder). With perfected cameras (such as described, for instance, in the second article quoted in footnote¹⁾) it is possible to work with angles

¹⁾ See, e.g., W. G. Burgers, The use of X-rays and cathode rays in chemical and metallographic investigations, Philips Techn. Rev. 5, 157-166, 1940; W. Parrish and E. Cisney, An improved X-ray diffraction camera, Philips Techn. Rev. 10, 157-167, 1948.

as small as 3° . When using $\text{CuK}\alpha$ rays these angles correspond to a separation of the lattice planes by about 15 \AA . With still smaller angles there is a troublesome background of scattered radiation.

This scattering gives rise to difficulties when it is desired to study "crystalline" substances whose lattice planes are spaced far apart, up to some hundreds of \AA , as is the case with proteins and other organic substances. The scattering angles of the reflected rays are then very small (less than one degree) and the intensity of the reflected rays from these organic substances is also extremely small, so much so that with a normal camera it is impossible to observe these reflections. A similar difficulty is encountered when it is desired to study the small-angle scattering of X-rays by a very fine powder as such. From this small-angle scattering, information can be obtained on the distribution of the grain sizes and on the mutual distances of the grains. The intensity of the scattered image decreases rapidly with increasing scattering, so that the image becomes of importance mainly with scattering angles smaller than 1° . But this image, which is still very weak, can be observed only if the scattering caused by the edges of the diaphragms and by the molecules of the air is suppressed as far as possible.

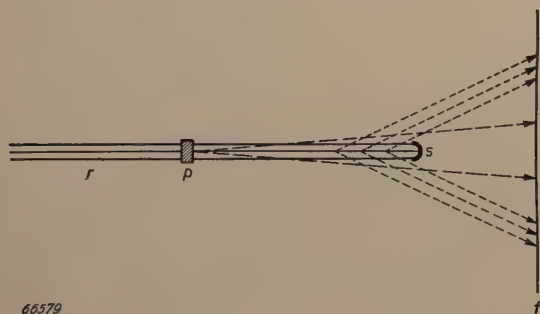


Fig. 1. The scattering of X-rays by the molecules of the air (see short-dash lines) takes place mainly over large angles. When the lead bead s intercepting the primary beam r is placed a few centimetres in front of the film f , practically none of the air-scattered rays fall within the area of small angles of diffraction, while the solid angle from which these scattered rays emerge is also small. The rays diffracted by the specimen p are represented by the long-dash lines.

For the investigation of this small-angle scattering special measures have to be taken, with the object of minimizing the intensity of the scattered radiation at such small angles of diffraction and increasing the intensity of the primary beam. In this case, it should be noted, the scattering by the air is of no consequence. The molecules of the air tend to scatter the X-rays at large angles to the direction of the primary beam. The concave

lead bead with which the primary beam is cut off for these investigations, so that it shall not blacken the film, is placed a few centimetres in front of the film, the rays scattered by the molecules of the air then striking the film only in the region of the larger scattering angles. At the same time the solid angle from which the scattered rays emerge is very small (*fig. 1*). Thus in our case the scattering at the edges of the diaphragms is the most important factor.

Various collimating systems

To reduce this edge scattering as far as possible, various systems have been developed. For instance, there is the system whereby a third slit diaphragm is added to the series of two diaphragms already

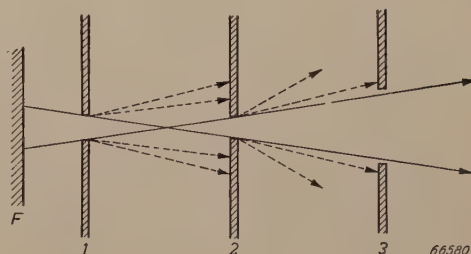


Fig. 2. Extended slit-diaphragm collimator. Of the rays emitted by the focus F only a narrow beam is allowed to pass through the apertures 1 and 2. The edges of these apertures cause scattering. Therefore a third diaphragm, 3, is added, usually placed at the same distance from 2 as that between 1 and 2. The width of this third aperture is such that the primary beam can just pass through without touching the edges. The rays scattered by the edges of aperture 2 are intercepted by the third diaphragm. The width of the apertures 1 and 2 is usually 100μ to 200μ , whilst the distances between the diaphragms are of the order of 10 to 20 cm.

mentioned (see *fig. 2*). Such a system requires very precise alignment (the slits are narrow and spaced far apart; the edges of the third slit must just be clear of the primary beam) and yields only a small intensity. The third diaphragm, moreover, makes the system longer and as a consequence the intensity at the specimen is much less still than that obtained with only two diaphragms.

A collimator based upon an entirely different principle is the planar crystal monochromator, the working of which is graphically explained in *fig. 3*. As its name implies, this collimator produces a monochromatic beam, this being an advantage on account of the different wavelengths produced by the X-ray tube. In addition to the $\text{CuK}\alpha$ rays a copper anticathode yields also $\text{CuK}\beta$ and $\text{CuK}\gamma$ rays, etc. Furthermore, there is continuous radiation, particularly at the smaller wavelengths.

It is only with monochromatic radiation that sharply defined diffraction lines can be obtained. With slit-diaphragm systems, both for small and for normal angles of diffraction, it is therefore always necessary to have a filter in the primary beam

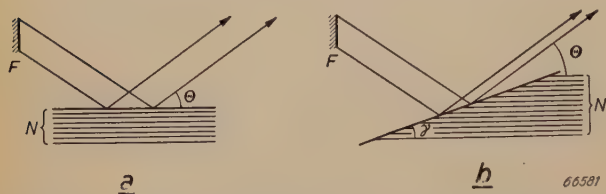


Fig. 3. Planar crystal monochromator. *a*) The rays falling from the focus *F* onto the crystal are reflected if their angle of incidence θ with the lattice planes *N* satisfies Bragg's law. All reflected rays are (theoretically) parallel; the width of the beam is determined by the width of the focus. *b*) Often the crystal is ground in such a way that the lattice planes make an angle γ with the surface which is only a little smaller than θ ; the reflected beam is then extremely narrow and a greater "brightness" is obtained. The depth to which the X-rays penetrate into the crystal is very small, of the order of a few lattice spacings. Here these distances are greatly exaggerated.

so as to make it monochromatic (i.e. to filter out the $K\beta$ rays, etc; the continuous radiation cannot be suppressed in this way). Naturally such a filter involves additional loss of intensity. With the crystal monochromator no filter is needed, and moreover a larger part of the beam emerging from the focus is utilized. Compared with the slit-diaphragm system, however, this crystal monochromator still does not give any appreciable increase in intensity, since about 80% of the intensity is lost in the reflection at the crystal.

By grinding the crystal concave and then bending it (bent crystal monochromator) a much larger part of the beam emerging from the focus can be made effective. But neither the planar nor the curved crystal monochromator is of any practical use for angles of diffraction smaller than about $10'$, owing to the irregularities in the crystal (mosaic structure).

Collimation by means of total reflection

In the Philips Laboratory at Eindhoven a collimator has been developed with which an X-ray beam with a divergence of about $5'$ can be produced in a very simple manner. This collimator yields an intensity about three times greater than that obtainable with a slit-diaphragm system for the same divergence. There is little scattering, the beam passed through is partly monochromatic and the construction is very sturdy. It is based upon the principle of the total reflection of X-rays.

For all solids the index of refraction for X-rays is less than 1 (contrary to that for light rays).

As a result, when passing from air to, say, glass, the X-rays can be totally reflected provided their angle of incidence is smaller than a certain critical angle η . As we shall see presently, the indices of refraction for X-rays by the various materials differ very little from unity, and this means that η is very small, so that only rays falling at a very small angle are reflected; all others are almost completely absorbed (disregarding here possible Bragg reflections).

Let us turn again to the simple slit-diaphragm system consisting of two apertures (fig. 4*a*). Of the rays emerging from point *A* at the anticathode of the X-ray tube only those which lie in the cone with boundary rays *p* and *q* are effectively used. Rays such as *p'* and *q'* passing through the first aperture do not reach the second one and are therefore lost. This accounts for the low intensity of the beam passing through a slit-diaphragm system. The effective aperture angle as seen from any point on the surface of the anti-cathode is determined by the width of the last slit and its distance from the focus.

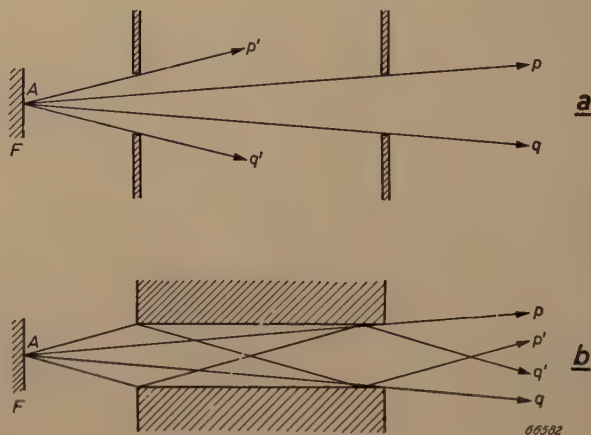


Fig. 4. *a*) With a slit-diaphragm system only those rays are passed through which lie in the solid angles formed by the points of the focus and the aperture of the last diaphragm, so that only part of the beam emitted by the focus is effectively used. *b*) When the collimator consists of two totally reflecting plates full use is made of all the rays lying in the solid angle formed by a point of the focus and the first aperture. (It is assumed that in the situation drawn total reflection is possible.)

The rays *p'* and *q'* can, however, be utilized by substituting for the two slit-diaphragms two polished glass plates (fig. 4*b*), from which these rays are totally reflected. (Here we are speaking of bilaterally limited beams and not of all-round limited beams; for the latter circular diaphragms would have to be used instead of slit diaphragms, and a cylindrical hollow tube instead of plates.)

Such a collimator has been described by Nähring²⁾. Thus the whole of the beam emerging from the point *A* passes through the collimator. Compared with a slit-diaphragm system with the same divergence the total-reflection collimator therefore gives a more intensive beam. It has yet another advantage, in that scattering of the rays at the edges of the slits is entirely avoided; since the rays are reflected from the plates there is no absorption and thus no scattering, not even at the edges at the end of the collimator.

The quality of this collimator, however, is bound to two limits. In the first place the angle of divergence of the incoming rays cannot be chosen haphazardly; only those rays from *A* striking the reflecting surface at an angle smaller than the critical angle η are passed through. The second limitation is that the divergence of the emerging beam cannot be made very small; it will amount to 2η , if the collimator is of sufficient length (a least just as long as a split-diaphragm collimator with the same divergence). If the collimator is made even longer this does not have any further influence upon the divergence of the emerging beam, as is the case with a slit-diaphragm collimator. Now the critical angles of the materials suitable for the collimator are in the order of $10'$ to $15'$, which means to say that with the total-reflection collimator beams cannot be obtained with a divergence less than $20'$ to $30'$. For investigations where this divergence is small enough this collimator offers the above-mentioned advantages over the slit-diaphragm system, but for the examination of proteins and suchlike, where a still smaller divergence (of a few minutes of arc) is required, the collimator is of no use.

However, by slightly modifying the construction it is possible to give the emerging beam a much smaller divergence, while still retaining the advantages of higher intensity and the minimum of scattering. This is achieved by mounting the polished plates at a small angle with respect to one another instead of parallel (fig. 5). Thus the collimator is given a wedge shape, the rays entering through the narrow opening and emerging through the wider one. Say that the angle made by the two plates to the plane of symmetry is β . A ray entering at an angle δ to this plane and striking the wall is reflected at an angle $\alpha - 2\beta$ to the plane (we are considering only the absolute value of the angle). After n reflections the angle made by the

ray with the plane of symmetry is

$$\alpha_n = \alpha - 2n\beta. \quad (2)$$

Thus the path followed by the ray will be more and more in the direction of the plane of symmetry. Since this applies to each ray, the divergence of the emerging beam will be smaller than that found with parallel plates. In this way a beam can be obtained with a divergence of about $5'$ without sacrificing any of its intensity.

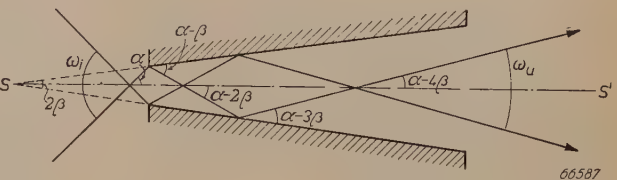


Fig. 5. The wedge collimator. Each of the two reflecting surfaces makes an angle β with the plane of symmetry ss' . A ray having an angle of incidence α with respect to this plane (and striking the wall) will be reflected if $\alpha < \eta + \beta$ (where η represents the critical angle of the material) and will be absorbed if $\alpha > \eta + \beta$. After having been reflected n times a ray entering with an angle of incidence α to the plane of symmetry will have only the angle $\alpha - 2n\beta$ to that plane. With a suitable choice of the dimensions of the wedge (determining n) and the angle of inclination β , the rays entering within the aperture angle $\omega_i = 2(\eta + \beta)$ will leave the wedge with a much smaller aperture angle ω_u . Here two rays are drawn which have the largest possible angle of incidence ($\alpha - \beta = \eta$). The angles are strongly exaggerated in this diagram.

Before considering further the path of the rays it will be useful to discuss briefly the phenomenon of total reflection of X-rays.

Total reflection of X-rays

The condition for total reflection is:

$$\sin i_1 > n_2/n_1, \quad (3)$$

where i_1 is the angle made between the incident ray and the normal to the boundary plane of the two media, n_1 and n_2 are the indices of refraction of the media ($n_1 > n_2$), whilst the direction followed by the ray is from medium 1 to medium 2. If the angle made with the surface is, say, $\varphi = 90^\circ - i_1$ then the condition for X-rays passing from air to a substance with an index of refraction $1 - \delta$ is:

$$\cos \varphi > 1 - \delta. \quad (4)$$

Thus the critical angle η , the largest angle at which total reflection is still possible, is given by:

$$\cos \eta = 1 - \delta,$$

or, since $\delta \ll 1$, to a very close approximation, by:

$$\eta = \sqrt{2\delta}. \quad (5)$$

²⁾ E. Nähring, Phys. Z. **31**, 401-418, 1930, in particular p. 405.

Since $\sqrt{2\delta}$ is very small, viz. less than 1.5×10^{-2} for all solids, total reflection of X-rays can be obtained only when they strike the surface at a very small angle ($\varphi < 50'$).

Following this reasoning one would expect that the coefficient of reflection R for rays having an incident angle smaller than the critical angle would be equal to unity, and for rays with an incident angle larger than η equal to zero.

Actually the variation in the coefficient of reflection is more gradual, as indicated in fig. 6.

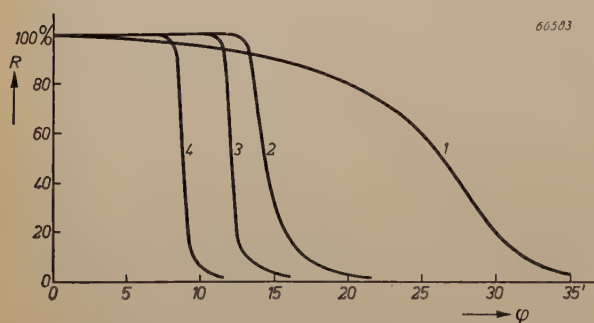


Fig. 6. Reflection curves for different materials. The reflection coefficient R is plotted against the angle of incidence φ (in minutes of arc). The lighter the material, the less it absorbs the X-rays and the smaller is the critical angle (the point where the curve bends), whilst also the curve is steeper. 1 silver, 2 glass not containing lead, 3 plastic material, 4 beryllium. These curves apply for $\text{CuK}\alpha$ rays of 1.54 \AA wavelength.

For angles of incidence slightly smaller than the critical angle the coefficient R is not yet equal to 1, whilst with angles larger than η reflection still takes place, though to a smaller extent. This is related to the absorbent properties of the material: the lower the absorption coefficient, the steeper are the curves and the smaller is the critical angle.

The curves drawn in fig. 6 apply for a wavelength of 1.54 \AA . For larger wavelengths the curves are displaced to the right and are less steep; the critical angle appears to be proportional to the wavelength.

Since the indices of refraction differ so little from unity it is not possible to make an X-ray lens. The focal length f of a simple refracting surface with radius r is of the order of r/δ , thus much greater than r . Therefore, in order to give a useful focal length an X-ray lens would have to be built up from a large number of components. If a material with large δ is to be chosen for this "lens" then that material would also have a large absorption coefficient. Consequently the total absorption in the X-ray lens is always far too great for the lens to be of any practical use.

The fact that ordinary light can be concentrated with lenses is due to the indices of refraction of solids for light being in general much higher than unity, while there are materials, such as glass, which practically speaking absorb no light at all.

Apart from collimation, which, as we have seen, can be brought about by total reflection, it has also been found possible to form an image by means of X-rays with the aid of this principle. With one or more spherically or elliptically ground mirrors an "X-ray lens" (actually it is not a lens) has indeed been made³⁾, and even an X-ray microscope has been designed, also with concave mirrors⁴⁾. Compared with an optical microscope such a microscope has the advantage that its resolving power is much greater, and compared with an electron microscope, the rays penetrate more deeply into the object being examined, so that also the internal parts of the object can be studied. However, it has not yet been possible to obtain a magnification of more than 30 times, and the aberrations are very considerable.

The path of the rays in the wedge collimator

We shall now consider briefly the relation between the directions of the incident rays and those of the emergent rays. The conditions to be satisfied if the incident rays entering through the narrow opening are to be reflected is that the angle of incidence at one of the reflecting planes must be smaller than η , or, that the angle α between the ray and the plane of symmetry (see fig. 5) shall be given by:

$$\alpha < \eta + \beta. \quad \dots \dots \dots (6)$$

The effective aperture angle of the incident rays is thus given by:

$$\omega_i = 2(\eta + \beta). \quad \dots \dots \dots (7)$$

Here it is assumed that the dimensions of the collimator are such that rays with an angle of incidence greater than $\eta + \beta$ cannot pass through the collimator without being reflected. If d_i is the width of the entrance and d_u that of the outlet, whilst l represents the length of the collimator, then it is assumed that the condition

$$\tan \alpha_0 = \frac{d_i + d_u}{2l} < \tan(\eta + \beta); \quad \dots \dots (8)$$

is satisfied; α_0 is the maximum angle of incidence at which a ray can pass through the collimator without being reflected.

To calculate the shape and divergence of the emerging beam an artifice is employed (fig. 7). By projecting mirrored images of the wedge from its reflecting surfaces it is possible to draw, for instance, instead of the zig-zag ray $APQRD$ the straight ray AD'' . Since in practice the angle β is only a few minutes of arc, the line $DCD'C'D''$ can be regarded as being straight, although it does not

³⁾ W. Ehrenberg, J. Opt. Soc. Am. **39**, 741-746, 1949, and P. Kirkpatrick, J. Opt. Soc. Am. **39**, 796-797, 1949.

⁴⁾ P. Kirkpatrick and A. V. Baez, J. Opt. Soc. Am. **38**, 766-774, 1948.

appear to be so in this diagram drawn on such an exaggerated scale.

The rays coming from *A* which pass through the wedge without being reflected at all have an angle of incidence *a* such that

$\tan a < \tan a_0 (9)$

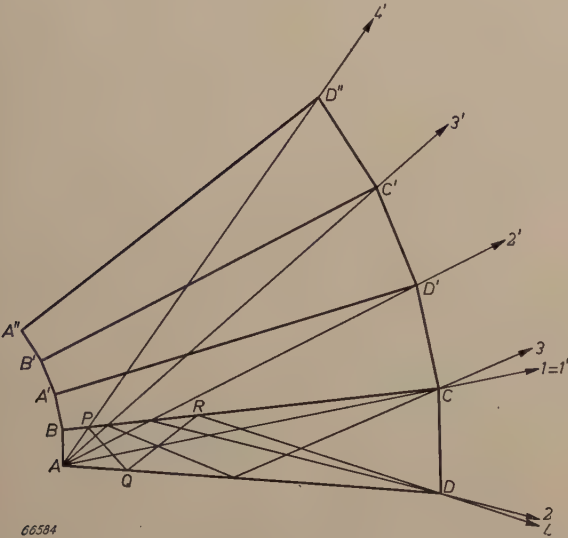


Fig. 7. Projection for calculating the path followed by the rays through the wedge.

A ray (passing through *A*) will be reflected once if its angle of incidence lies between *a*₀ and *a*₁, the latter being given by:

$\tan a_1 = \frac{d_i + 3d_u}{2l} , (10)$

as follows directly from fig. 7. So we may continue and ultimately find that a ray passing through the point *A* will be reflected *n* times if its angle of incidence lies between *a*_{*n*-1} and *a*_{*n*}, where *a*_{*n*} is given by

$\tan a_n = \frac{d_i + (2n + 1) d_u}{2l} (11)$

Since all angles are very small, in these formulae the tangents may be replaced by the angles.

In this way it is also possible to follow the rays not entering at *A* but at some other point in the entrance opening. By calculating the angle of emergence of all these rays with the aid of formula (2) we arrive at the shape and divergence of the emerging beam.

Calculations show that the intensity of the emergent rays is practically the same in any permissible direction of emergence, whilst the distribution of the intensity over the whole width of the wider outlet is uniform. The outlet can easily be made 100 to 200 μ wide, thus giving it a width

comparable to that of the slits in a diaphragm system. Although the width of the entrance to the wedge collimator is much smaller than that in a slit-diaphragm system, the intensity of the emergent beam is much greater, owing to the larger solid angle ω_i for the incidence of the rays; in the slit-diaphragm system this angle is equal to the divergence of the emerging beam, whereas in the wedge collimator it is much larger.

The gain in intensity compared with a slit-diaphragm system yielding a beam of the same width and divergence amounts to a factor of 3 to 5.

The wedge also functions somewhat as a monochromator. The effective aperture angle for the incident rays is given by (7), and since β is always much smaller than η it may be assumed, without making any great error, that the effective aperture angle is proportional to the critical angle. Now for all materials the rule is that η is proportional to λ , and so the aperture angle will also be proportional to λ . This is just what is desired: the continuous part of the radiation emitted by the focus, which part has mainly shorter wavelengths and cannot be intercepted by a filter, is attenuated with respect to the monochromatic component, the ratio of this attenuation being equal to that of the wavelengths, i.e. for normal tube voltages a factor of 2 to 3.

The construction of the wedge is determined by the critical angle of the material and the divergence desired. As will presently be shown, it is preferable to choose a material with the steepest possible reflection curve, thus also with a small critical angle. Common window glass with $\eta = 14.5'$ (curve 2 in fig. 6) answers very well; still better would be an organic material (such as plastics) containing only light atoms, but for our purpose this is not sufficiently stable in shape.

As an example the dimensions are given of a wedge collimator made of window glass in this laboratory, the construction of which is outlined in fig. 8. The length is 120 mm, the entrance is

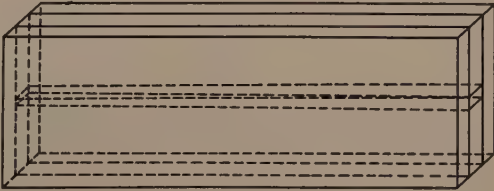


Fig. 8. Construction of the wedge collimator. The reflecting plates are placed at the precise distances and with the precise inclination with the aid of gauges and then firmly cemented between two other plates. The whole is very sturdy and easy to make.

9 μ wide and the outlet 140 μ wide. With CuK α radiation a beam can be obtained with a divergence of 4.8 minutes of arc. According to formula (7) the effective aperture angle for the incident rays is 33' (since $\beta = 2'$).

Around the emerging primary beam with a divergence of approximately 5' there appears a "halo" of about 25' diameter. This halo is not to be ascribed to scattering by diaphragm edges such as arises with a slit-diaphragm collimator, and it is therefore of much lower intensity. It arises from other causes of "scattering".

"Scattering" in the wedge collimator

There are three causes of this scattering in the wedge collimator.

1) As a result of the irregularities in the surfaces of the glass plates some scattering is bound to occur. Assuming that by thorough polishing the differences in thickness of the plates can be reduced to about 20 Å, it may be estimated that for the small angles of incidence with which we are dealing here the scattering due to such irregularities will not be of much consequence. Owing to this effect the scattering halo might have an aperture angle of about 15'; this does not, therefore, account for the 25' mentioned above.

2) Owing to the "tail" of the reflection curve also rays with angles of incidence larger than η are reflected, thus giving the emergent beam a greater divergence. This effect, of course, is not scattering in the real sense of the word, being rather a regular reflection, but here it is considered as scattering because it likewise contributes towards the formation of a halo around the primary beam. The larger the number of times the ray is reflected, the smaller is the angle the ray makes with one of the polished surfaces. Consequently a ray striking one of the planes for the first time at an angle $\varphi > \eta$, and thereby being reflected, even if for only a small percentage, will be much better reflected the second time because then the angle of incidence will be nearer η , or it may perhaps have become smaller than η . At the next reflection the ray will be still better reflected.

The decrease of the angle of incidence is determined by formula (2). When, as is the case in practice, φ is only a few minutes of arc, all rays with an incident angle $\varphi > \eta$ will be considerably weakened when passing through the collimator, so that with $\varphi > 1.5 \eta$ the intensity of the emergent rays will be about one million times smaller than that of the incident rays, as a rough calculation shows. Rays with an incident angle φ between η and 1.5η will emerge partly within the angle of divergence of the rays with an incident angle $\varphi < \eta$. This contribution towards the intensity of the emergent beam is an advantage, albeit a very slight one. Part of these rays, however, have a larger divergence and give rise to a halo around the emergent beam. It appears that owing to the "tail" of the reflection curve the aperture angle of this halo will amount to about 8'.

3) The polishing agent, which contains iron, leaves some iron atoms behind in the uppermost layer of the surface, and when these atoms are excited by the incident X-rays they themselves radiate in all directions, thus giving rise to "secondary X-rays". This effect can be made quite perceptible with the aid of various filters introduced in the beam. It is

possible that this accounts for the divergence of 25' for the scattering halo. This halo could be reduced by polishing, for instance, with aluminium oxide.

Use of the wedge collimator

The specimen to be examined is placed immediately behind the wide opening, and the film 10 to 20 cm behind that. A nickel filter is placed in front of the wedge so as to suppress the CuK β rays.

As an example of a photographic recording *fig. 9* gives a diffraction pattern of wet collagen. The

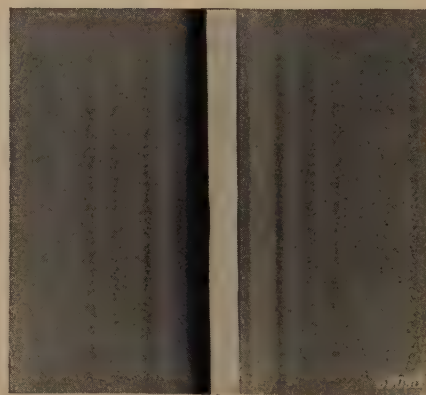


Fig. 9. Diffraction pattern of a wet collagen specimen 1 mm thick, taken with the aid of the wedge collimator (scale of the reproduction 3:1). Focus 1.2 mm \times 1.2 mm, tube voltage 30 kV, tube current 30 mA, distance object to film 200 mm, exposure 20 hours. The spacing in wet collagen is 660 Å and the reflections of the second and higher orders are quite perceptible (mutual distance 8'). The reflection of the first order is rendered invisible by the scattering background. The primary beam was intercepted with the aid of a lead bead.

distance of the most important lattice planes of this protein is 660 Å, and thus with CuK α X-rays reflections are found with 8' between their successive orders. The reflections of the second and higher orders are quite perceptible. The reflection of the first order (angle of diffraction 8') is not visible

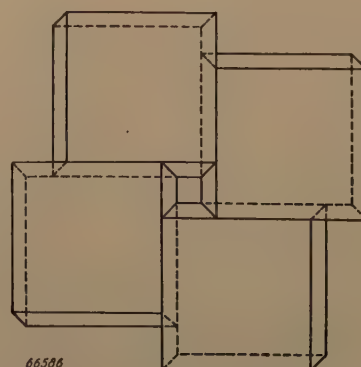


Fig. 10. Sketch representing the construction of a collimator according to the principle described here and limiting the X-ray beam on all sides (seen from the end with the wide opening),

owing to the background of scattered radiation. The exposure for this recording was 20 hours. With a slit-diaphragm system of the same divergence an exposure of 75 hours would have been required.

A collimator according to the principle described here can also be constructed for limiting a beam on all sides. This is of importance for many applications, e.g. for examining single crystals in different directions. The principle of the construction is represented in *fig. 10*. Theoretically, with such a collimator an all-round limited beam can be obtained with an intensity 10 to 30 times greater than a beam of the same diameter and divergence limited by a number of diaphragms with small aperture.

Summary. For studying the structure of certain substances it is necessary to be able to measure the diffraction of X-rays over very small angles ($<1^\circ$). This requires an X-ray beam with very small divergence, of a few minutes of arc. To satisfy this requirement and at the same time obtain sufficient intensity of the beam and a sufficiently weak scattering background, special collimators have been developed, which are briefly described. The authors have constructed a collimator which is very sturdy and quite simple to make. This collimator, with which a divergence of about $5'$ can be obtained, is based upon the principle of the total reflection of X-rays when they strike a polished surface at a very small angle. By means of a wedge-shaped arrangement of the reflecting planes, which can be made, for instance, from window glass, a beam of about $30'$ divergence can be reduced entirely to the much smaller divergence mentioned, thereby producing a relatively intense beam. Consideration is given to the causes of the scattering occurring, which is found to arise to a smaller extent than is the case with other types of collimators. Finally an example is given of the results that can be achieved with this collimator.

MEASURING RAPIDLY FLUCTUATING GAS TEMPERATURES

by B. H. SCHULTZ.

621.317.39:536.5:536.717:621.4/.5

Petrol engines, diesel engines and suchlike are being used by the million and yet the processes taking place inside them are still not known in great detail. For engines which are still in course of development, such as gas turbines, air engines, etc., this obviously applies all the more. Perhaps the most important detail in respect to which a deeper insight is desired in all these cases is the variation of the gas temperature in the engine cylinders during a cycle of the work process. The method of measuring described here promises to broaden considerably the knowledge we have about this problem.

Fluctuating temperatures in thermal engines

One of the deciding factors for the work process taking place in any kind of thermal engine, be it a steam or a petrol engine, a jet or an air engine, is the temperature of the gases taking part in the process. This may be explained by an example. In the case of the air engine ¹⁾ the gas (in this case air) is presumed to have a constant and uniform high temperature T_w in the hot space and a constant and uniform low temperature T_k in the cold space. Energy is obtained by causing the gas in the hot space to expand and then transferring it via the regenerator to the cold space and compressing it, after which it is transferred back to the hot space via the regenerator, the cycle then starting again. The efficiency of the process in the ideal case (when there are no losses whatever) is:

$$\eta = \frac{T_w - T_k}{T_w} \dots \dots \dots (1)$$

Actually these temperatures T_w and T_k are not exactly constant with time, and the gas temperature also varies from point to point. The walls of the hot and cold spaces are indeed at constant temperatures, but in the course of expansion the gas in the hot space is slightly cooled, especially in the central part of the gas volume, since the heat exchange with the wall cannot be complete in the very short time available (in the case of an engine running at 3000 r.p.m. the time for each expansion is only about 0.01 second). Much the same applies for the gas in the cold space during compression. Thus at any point in the gas space there will be rapid periodical fluctuations of the gas temperature. Equation (1) is then no longer valid. The power and efficiency then have to be calculated from the periodical variation of pressure as influenced by the fluctuating gas temperatures.

In principle it is possible to calculate these fluctuating gas temperatures from constructional data and the working conditions of the engine, not only of that taken as an example here but also for other engines, but when making such calculations one is faced with considerable difficulties. A method has therefore been developed by means of which fluctuating temperatures can be measured quickly and with the desired accuracy. The novelty of the method to be described here is not to be sought in its principles but in the details of application imposed by the exceptional requirements.

Principle of the method of measurement

The method is based on the use of a resistance thermometer which consists of a thin wire of a suitable material introduced into the gas at the point where the temperature is to be measured, forming, one of the four arms of a bridge. Any variation in the temperature of the filament causes a variation in its resistance, the bridge is thrown out of balance and the diagonal voltage of the bridge serves as a measure for the change in temperature. By applying this voltage to a cathode-ray oscillograph it is possible to record and measure also rapid fluctuations in temperature. In our case, where the fluctuations occur periodically with a cycle equal to the duration of one revolution of the engine (twice that in the case of a four-stroke engine), it is most convenient to synchronize the time base of the oscillograph with the engine. Thus a stationary picture of the temperature variation is obtained on the screen, in so far as the temperature fluctuations occurring are sufficiently reproducible.

Fig. 1 is a photograph of the measuring set-up. Further on the reader will find reproductions of some oscillograms obtained with this set-up, with explanations in the subscript.

¹⁾ See, e.g., H. Rinia and F. K. du Pré, Air engines, Philips Techn. Rev. 8, 129-136, 1946.

Thermal inertia of the thermometer wire

Owing to its thermal inertia the thermometer wire will respond to the gas temperature with the required accuracy only when the engine is running at a sufficiently low speed. As the speed increases, the temperature of the wire lags more and more behind that of the gas, and to a first approximation the situation might be regarded as if there were only a delay in the wire temperature, the form and

from the gas to the wire in the short time dt is proportional to the surface of the wire and to the difference between the temperatures T_g and T_d of the gas and the wire respectively.

From eq. (3) it is easy to derive eq. (2), since dQ is the heat absorbed by the wire in the time dt , thus being equal to the product of thermal capacity and rise in temperature of the wire ²⁾: WdT_d . Hence eq. (3) may be written in the form:

$$T_g = T_d + \frac{W}{aF} \frac{dT_d}{dt} \dots \dots \dots (4)$$

Denoting W/aF by τ , the right-hand term may be regarded

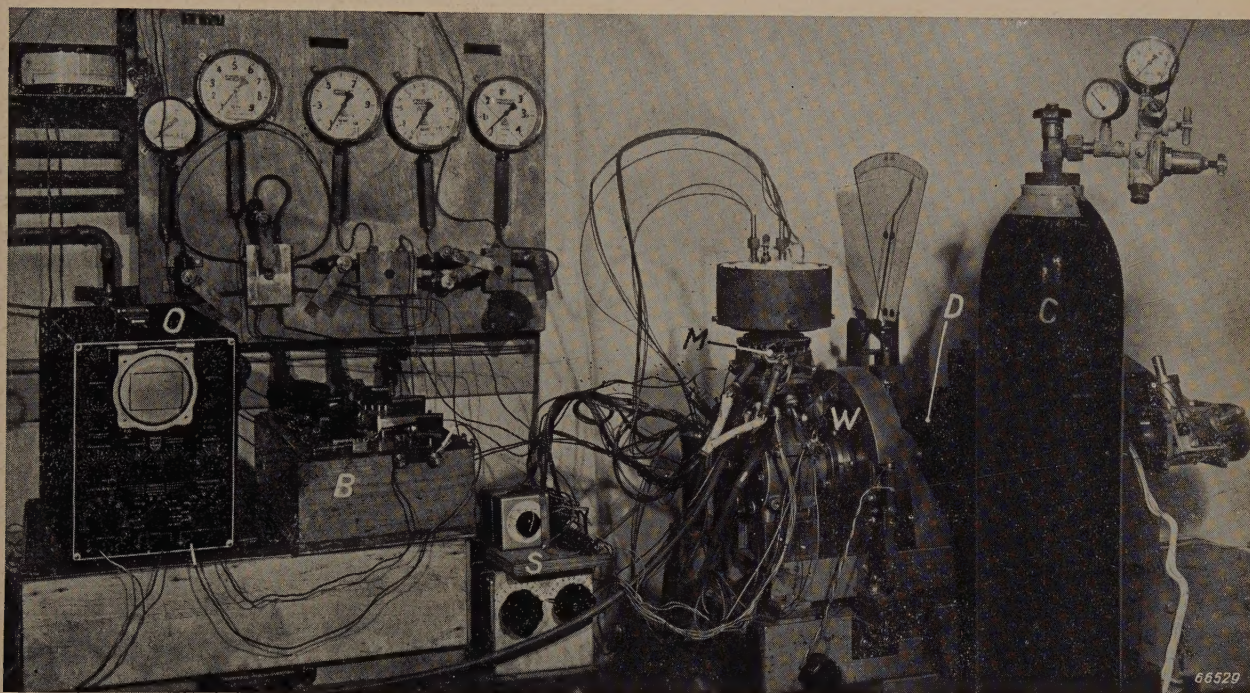


Fig. 1. The complete measuring set-up. M is an air engine in which thermometer wires have been introduced at various places. B are the variable resistors of the bridge circuit. O is the cathode-ray oscillograph. S are switches for connecting one of the wires as required. In this set-up there are coupled to the engine a flywheel W (to which is attached the interrupting mechanism referred to later on) and a brake dynamo D . The pressure inside the engine is adjusted with the aid of the high-pressure cylinder C . The manometers at the top of the picture on the left serve for various static pressure measurements.

amplitude of the temperature variation of the wire still being precisely the same as that of the gas. The time lag τ is given by:

$$\tau = \frac{W}{aF} \dots \dots \dots (2)$$

where W is the thermal capacity of the wire, F its surface area and a the coefficient of heat transfer between gas and wire, which is determined mainly by the geometrical configuration (thickness of the wire).

The coefficient a is defined by the equation:

$$dQ = aF(T_g - T_d)dt, \dots \dots \dots (3)$$

which expresses that the amount of heat dQ passing over

as the MacLaurin development of $T_d(t + \tau)$ broken off after the second term, so that eq. (4) gives expression to the fact that at the instant t the gas temperature T_g is approximately equal to what the temperature of the wire T_d will be at the instant $t + \tau$. Thus, with respect to T_g , T_d lags behind by a period of time τ , which is given by eq. (2).

Owing to the fact that the exact MacLaurin development for $T_d(t + \tau)$ contains terms of the second and higher derivatives which do not occur in the right-hand term of eq. (4), the wire cannot exactly follow sharp changes in the gas temperature but shows rounded-off temperature curves. Such a distortion, however, is of less practical importance than another which will now be dealt with. The remedy that will be discussed greatly reduces both distortions simultaneously.

²⁾ At least if it is permitted to disregard the heat transfer between the wire and the surroundings through radiation or through conduction via the extremities; see below.

When the time lag τ is known, the temperature curve as recorded could be corrected accordingly. Actually, however, this description is still too rough, for account must be taken of the fact that τ is also somewhat dependent upon the velocity and the pressure of the gas flow, and that α , and with it τ , therefore varies during the stroke of the engine. We could take an "average delay", but the varying value of τ naturally implies a certain distortion of the temperature curve which is not easy to correct. It is preferable, therefore, to try to make τ so small that the correction can be entirely ignored, even at the highest engine speeds.

To attain this object it was necessary to make the thermometer wire extremely thin. It is easy to understand that the thinner the wire the better it is able to respond to changes in the gas temperature, but it is well to define this more precisely, because the manner in which α depends upon the thickness of the wire d plays an important part in this case. With a given length of wire the surface area F is proportional to d , whilst the volume and thus the thermal capacity W is proportional to d^2 , so that owing to these two factors in eq. (2) the lag τ already decreases proportionately with d . Moreover, as empirically established, the coefficient of heat transfer α changes in proportion to $d^{-0.7}$, so that with decreasing thickness of the wire the delay time is reduced at an even higher rate, viz. in proportion to $d^{1.7}$.

For our thermometer we mostly use Wollaston wire, this being made by coating fine platinum wire with silver and drawing it down further so that the platinum core is reduced, in our case, to a thickness of from 1.5 to 2 microns; after the wire has been mounted the silver is dissolved off by etching, so that only the platinum wire of the thickness mentioned remains. We have also frequently used tungsten wire reduced to such a fine thickness by etching, and which is strong enough to be handled in this form. With these extremely fine wires the coefficient of heat transfer α reaches the exceptionally high value of about 0.3 cal/cm² degree sec (in the case of normal configurations where a gas gives off heat to a wall the coefficients α are usually ³⁾ in the order of 10⁻³ cal/cm² degree sec). The time lag τ is from 1 to 1.5 \times 10⁻⁴ sec, corresponding, at an engine speed of 3000 r.p.m., to a phase angle of from 1.5 to 2.5 degrees, a correction which in our case may reasonably be ignored, so that *a fortiori* the variations of α are of no consequence.

³⁾ The difference agrees well with the power -0.7 in the relation between α and d . It has been directly proved by experiments that for our extremely fine filaments the $d^{-0.7}$ law is still valid.

It will now be understood why we do not work with thermoelements, which are generally used for measuring wall temperatures of cylinders and the like. The joint of a thermoelement, suitably made, may well allow of a somewhat more accurate localization of the measuring point than is possible with a resistance wire, but the thermal inertia of the joint cannot be reduced to the same extent as that of our wires.

Incidentally it may be pointed out that the great increase of α with decreasing thickness of the wire was one of the decisive factors in the development of the modern gas-filled incandescent lamps. The long and thin tungsten wires answering to the desired voltage and the desired output involve in themselves large losses in the form of heat given off to the gas. By coiling the filament it is made equal, as far as heat transfer is concerned, to a wire of a thickness corresponding to the diameter of the coil, so that the thermal losses are reduced to a fraction of what they were originally.

With such fine metal wires the temperature is equalized over the whole cross section of the wire so quickly (within about 10⁻⁸ sec) that even with the highest engine speeds the temperature distribution over the cross section may be regarded as uniform. Thus it is possible to calibrate the set-up statically. By giving the resistance wires a length at least 2000 times their thickness (i.e. in our case at least 4 mm) it is possible to reduce also the measuring error due to lack of uniformity in the distribution of the temperature over the length of the wire (owing to conduction of heat to or from the electrodes between which the wire is stretched), so that for our purpose it may be ignored.

The temperature of the wire could be influenced to a certain extent also by the (constant) temperature of the walls of the cylinder owing to heat exchange through radiation, but with the intensive exchange of heat through convection between the flow of gas and the very thin wires this effect becomes negligible; there is, therefore, no need to counteract this by screening, not even at the highest temperatures at which measurements have been taken (about 600 °C).

The bridge current must, of course, be sufficiently low to avoid any perceptible heating of the resistance wires; a current of a few milliamps was found to be quite permissible.

Calibration of the oscillograms

Once an oscillogram of the temperature fluctuations at a point in the engine has been recorded, this has to be calibrated. That is to say, from the vertical deflections the absolute temperature values have to be deduced and each point of the curve has to be correlated to the right phase of the cyclic process. These two operations are carried out in one step, in the following way.

Before being mounted in the engine each resistance wire is statically calibrated with the aid of an adjustable furnace: the bridge is balanced for a number of different, known, wire temperatures by adjusting one of the bridge resistors, so that it is known what gas temperature corresponds to each value of this resistance. Mounted on the shaft of the engine under test is a mechanical interrupter (one working with a photocell can also be used) — see fig. 1 — which in one complete revolution, thus in one cycle of the oscillogram, cuts out the bridge current twice, or if necessary more frequently, for a brief moment and thus momentarily reduces all bridge voltages to zero. As a result, in the oscillogram there are two discontinuities where the spot returns for a moment to the level corresponding to the zero point of the diagonal voltage of the bridge (for the bridge voltage to drop quick enough

In fig. 2 we have two temperature curves recorded in the manner described. These have been obtained from measurements taken on an experimental single-cylinder air engine. Fig. 3 is a schematic cross-sectional drawing of the engine and shows the various points where thermometer wires were mounted. The measuring points from which the oscillograms *a* and *b* were obtained are denoted by the corresponding letters. From the oscillograms the temperature can be derived with an accuracy of about 1°C and the phase with an accuracy of a few angular degrees, which was more than sufficient for our purpose. From the shape of the curves important conclusions could be drawn as to the instantaneous value of the heat transfer at various points in the heat exchangers, the mixing of the gas in the cylinder and the heat transfer between the gas and the walls of the cylinder.

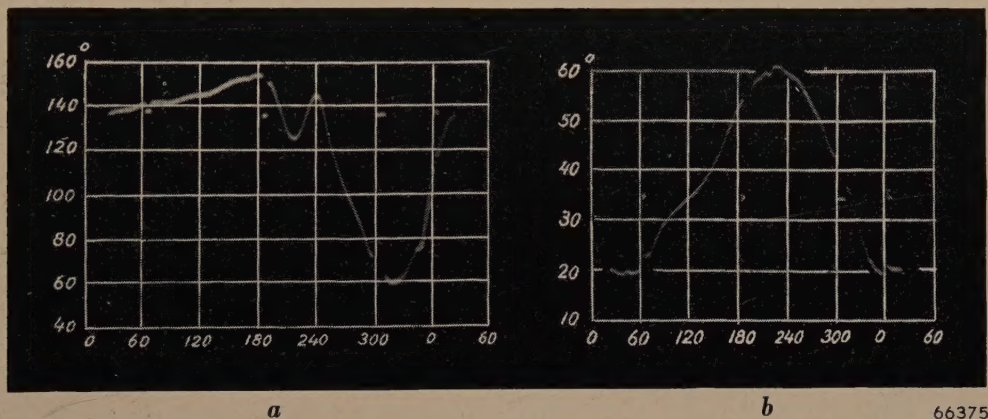


Fig. 2. Oscillograms showing the variations of the gas temperature at two places in an air engine; these points (*a*, *b*) are indicated in fig. 3. The phase angle 0° corresponds to the moment at which the volume of the hot cylinder space is greatest.

to zero the capacitances of the arms of the bridge must be sufficiently low). The bridge balance is varied by adjusting the resistance previously mentioned in such a way that the curve in the oscillogram is raised (or lowered) just sufficiently to cause one of the discontinuities to disappear. At that moment of interruption — the phase of which in the cyclic process is exactly known — the diagonal voltage is then equal to zero also without the interruption, and from the adjustment required the temperature at that instant can be found. After having applied this procedure to both the jumps we have sufficient data from which to plot a complete phase and temperature scale in the oscillogram — at least provided the deflection of the spot is a linear function of the bridge voltage; if this is not the case within the desired accuracy of a few percent then we must use more interruptions of the curve.

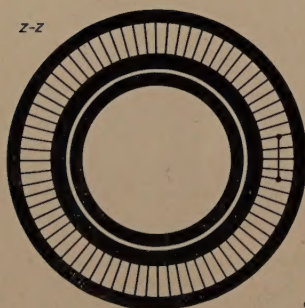
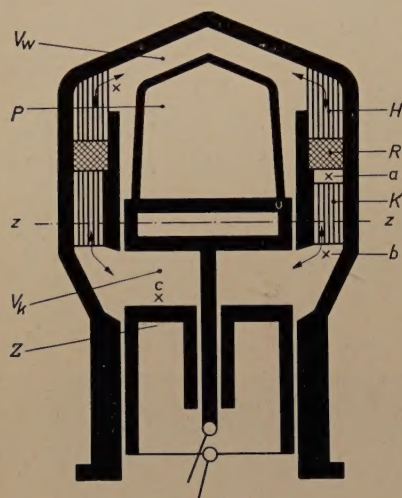
Mounting the wires

In conclusion something may be said about the manner in which the wires were mounted in the engine. Each wire is affixed to its electrodes with a minimum quantity of solder; for measuring points, where high gas temperatures occur the soldering is done with silver, using a burner with a very fine flame. The electrodes have to be insulated and the opening in the cylinder wall through which they are passed has to be sealed air-tight. A successful method is to use small metal tubes of 1.6 mm internal diameter containing an electrode rod of 1 mm diameter insulated with magnesium oxide pressed in between. Two of such tubes are soldered into borings in a flange. The Wollaston wire is then mounted between the electrodes, etched to remove the silver coating and then statically calibrated in the manner described. Finally the flange with the

electrodes is fixed air-tight over an opening made in the wall of the cylinder (see *fig. 4*).

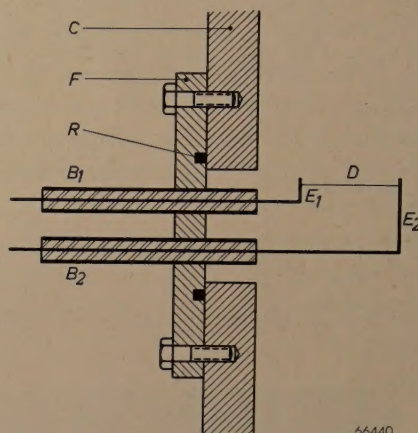
It may seem astonishing that it should be possible in practice to work with wires no more than 1.5 to 2 microns thick, which are almost invisible to the naked eye. What is perhaps most surprising, however, is that it has also been possible to measure with such a wire the fluctuating temperature of

the gas in the thin layer near the surface of the moving piston. For this test the filament was fixed to the piston and thus moved up and down with it (*c* in *fig. 3*). The electrodes were passed through the body of the piston and connected to the bridge circuit quite simply by means of flexible, coiled wires connected to fixed, insulated electrodes outside the cylinder. This apparently so fragile



66439

Fig. 3. Simplified cross-sectional drawing of a single-cylinder air engine with displacer piston (see *fig. 7* in the article by F. L. van Weenen, *The construction of the Philips air engine*, *Philips Techn. Rev.* 9, 125-134, 1947). V_w is the hot space, V_k the cold space, H the heater, R the regenerator, K the cooler, Z the working piston, P the displacer piston (which transfers the air expanded in V_w via $H-R-K$ to V_k and after compression in V_k back again, via $K-R-H$, to V_w). Thermometer wires were mounted at the points marked with a cross. The orientation of the filaments a and b is given in the horizontal cross section ($z-z$). These wires yielded the oscillograms of the variation of the gas temperature as reproduced in *figs 2a* and *b*. In order to mount the wire a it was necessary to remove a small portion of the cooler, so that the dead space was thereby slightly increased.



66440

Fig. 4. Insertion of the electrodes E_1 and E_2 for a thermometer wire (D) through the wall (C) of the engine cylinder. B_1 , B_2 metal tubes with internal insulation of magnesium oxide, F flange, R gasket to make an air-tight seal.

system proved to be of sufficient mechanical strength to stand up against several millions of cycles of the engine. Only now and then did a thermometer wire come to an untimely end owing to minutely small droplets of lubricating oil floating in the cylinder being deposited on the wire and destroying it through their inertia.

Summary. In the cylinders of high-speed thermal engines, where one cycle lasts only a few hundredths of a second, very rapid fluctuations take place in the temperature of the gas, which fluctuations are of great importance for the functioning of the engine but are almost impossible to calculate. It has been found a practicable possibility to measure such rapid temperature fluctuations with the aid of an extremely fine metal wire connected as a resistance thermometer to an electric bridge. This wire, of a thickness of from 1.5 to 2 microns, responds to the variations in the temperature of the gas with a time-lag of only 1 to 1.5×10^{-4} second. The fluctuations are recorded with the aid of an oscillograph: the oscillograms can be calibrated in a simple manner. The error in the measurement of the temperature is about 1°C , while the error in the phase angle is at most a few degrees. As an example it is described how measurements were made on an experimental single-cylinder air engine.